ABSTRACT

USING SAMPLE VALIDATION IN AUDITING

A. Ochigbo, M.S.
Division of Statistics
Northern Illinois University, 2014
Dr. Alan M. Polansky, Director

A state government agency wants to determine if a sample verification procedure, used in auditing accounts to detect fraudulent charges, reduces the error of estimating the unknown mean amount of fraud per transaction. The procedure consists of checking whether the covariates of an audit variable, whose mean is known, is within a $100(1-\alpha)\%$ confidence interval computed on the observed sample. In this study we use computer-based simulations to explore the effect that the procedure has on the error in estimating the mean. We concentrate on the bivariate normal distribution and on a normal bivariate normal mixture. Numerical results are presented that compare the estimated error for estimating the unknown population mean using the implementation of the sample acceptance algorithm and the standard method based on simple random sampling. The proposed method reduces the estimation error for multivariate normal distribution but can have the opposite effect for the non-normal distribution. Another approach based on the well-known method of control variates produces similar results without the need to reject potential samples.

NORTHERN ILLINOIS UNIVERSITY
DE KALB, ILLINOIS

DECEMBER 2014

USING SAMPLE VALIDATION IN AUDITING

BY

A. OCHIGBO

A THESIS SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF SCIENCES

DEPARTMENT OF MATHEMATICAL SCIENCE

Thesis Director:
Dr. Alan M. Polansky

ACKNOWLEDGEMENTS

My heartfelt thanks goes to the faculty and the members of the staff of the Division of Statistics for their support and kindness. I am especially grateful to Professor Alan Polansky for his love, support, patience, and for ensuring that this project is a success. Thank you.

DEDICATION

To all who stood by me in this struggle. To my family and friends.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Financial auditing entails the verification of the financial statements of a firm or organization. The purpose of a financial audit is to assess the correctness of an organization's financial statements. A financial audit relates to information integrity and reliability. It often entails detailed and substantive testing.

One reason for a financial audit is to ensure that the firm is not engaging in deceptive practices. Another reason for a financial audit is to assess weaknesses in the internal control of accounting and financial reporting practices. Fraud is endemic in the financial reporting community. For example in 2001, Enron, an American energy conglomerate based in Houston, Texas, was discovered to have hidden important financial details from both stakeholders and the banks[12]. Enron filed for bankruptcy in one the the world's largest financial scandals of all time [13]. Also, in 2005, the United States Government Accountability Office (GOA) identified weaknesses in the fiscal year 2005 consolidated financial statements (CFS) audit. The GOA reported that the U.S. government did not have adequate systems, controls, and procedures to properly prepare the CFS. Such weaknesses impairs the ability for financial audit reporting to be consistent with established accounting principles [10]. For example, fraud costs millions of dollars every year to the Medicare program through various schemes by individuals, doctors or suppliers, and groups of

individuals. Fraud schemes include health care providers billing Medicare for services never received, doctors rendering services that are not medically necessary, and much more. Recently the U.S. Department of Health and Human Services released an audit report of the State Medicaid Fraud Control Unit (MFCU) for prosecuting patient abuse, and detecting and deterring fraud. The audit concludes that government's return on investment (ROI) is about nine to one for operations of the MFCU [8].

Further, a 2006 study by the Association of Certified Fraud Examiners (ACFE) estimates that about 6% of revenues of U.S businesses are lost to fraud. The ACFE also reported that small businesses are the most vulnerable to fraud (e.g., cash theft). Further, fraud is not confined to U.S businesses alone. Fraud at Ahold (The Netherlands), Parmalat (Italy), and Addeco (Switzerland) reveal that fraud is widespread and not localized to any particular region [2].

In auditing, the extensive use of sampling in making decisions is often employed. Therefore, the need for the sound application of statistical techniques in making decisions cannot be overemphasized. A consequence of not applying the correct statistical techniques, or of applying them haphazardly, can result in large monetary losses due to an inaccurate assessment of the financial situation of a company or other agency. In fact, the potential monetary losses accrued in auditing due to poor statistical techniques are gravely high [7], and many auditing procedures are primarily subjective. Poor auditing techniques may also expose a company or agency to legal liabilities.

As an example, an agency of the state of Illinois audits the accounts of certain vendors when there is some potential evidence of wasteful spending, abusive expenditures, or fraud such as overcharges and charges for non-existent services. The agency needs to know how much the vendor should be held liable for when such

wrong-doing is uncovered. The number of transactions is typically very high and an audit of each transaction is not possible. So a sample of the transactions is taken from the population of transactions with the vendor, along with some covariates whose exact means (over the population of all of the transactions) are known. For example, in Medicaid transactions there are often charges associated with transporting patients who qualify to their medical appointments. Such a transaction will not only include the amount charged, but may also include the distance traveled to the facility where the patient has the appointment, as well as other covariates such as which facility is being used, and patient information. Considering the entire database as a population of records, the means of the covariates are known, and can be used to assess the representativeness of the sample from the population of records.

One method employed by the state agency is to check to see if the true known mean of each covariate is within a $100(1-\alpha)\%$ confidence interval for the mean computed on the sample. If this occurs, then the sample is accepted. Otherwise, another sample is taken and the process starts again. The crucial question is whether the acceptance sampling scheme increases the accuracy of the estimate of the vendor's liability. This is an important question because issues of liability often end with litigation, and the validity of the estimates will need to be able to withstand potential legal challenges.

Sampling is not an uncommon practice in financial auditing. For example, consider an audit of the accounts of a bureau of public assistance with a large caseload of those in need of assistance. The purpose of the audit is to obtain an estimate of the distribution of assistance payments by amounts. Decisions have to be made quickly, but the large number of completed surveys mitigates against an intelligent assessment of records. Sampling often can provide such information quickly with the

necessary accuracy at a reasonable cost. Considering another example, most state departments of revenue use sampling in sales and use tax audits when the volume of records to be examined is enormous. The data collected from taxpayers are first verified before the auditor determines which accounts are selected for review. The audit population (sampling frame) is defined based on records submitted by the taxpayer and a summary is prepared that reconciles the number of transactions in the audit population with the data collected from taxpayers. The audit population is then stratified, a method which improves the efficiency of the statistical estimators used in the audit while preserving validity. Credits and liabilities are then projected accurately and efficiently based on the samples. Many additional examples can be seen in both government and business, where decisions have to be made quickly and accurately on the basis of information contained in voluminous records. Sampling can provide relevant information quickly and accurately with relatively low monetary rates.

The above examples lead to the following general conclusions:

- In each example the volume of transactions is large. If the entire number of records is small, then the application of statistical sampling may not be required.

- Sampling was of a recurring nature in the examples above. Hence, changes in the population of records of transactions were gradual with respect to time, so a sampling plan could be instituted with periodic modifications.

- Records of transactions should be such that samples can be selected with ease. These are usually now held in a large computer database so that sampling is a simple matter.

- It is important to have some prior knowledge of the accounts to be estimated from the sample, and to compare the sample estimate with the known value.
- Auditors often encounter voluminous records of transactions, so their sampling is usually of a recurring nature, hence, the need for the application of statistical sampling techniques.

It is important to note that most audits involve sampling because the auditor cannot examine all of the transactions of the population accounts. Moreover, the auditor must reach a conclusion about the accuracy of the accounts examined. The purpose of implementing sampling plans is to accurately estimate the amount of misstatement or error in an account. If the sample is not representative of the population of transactions available to the auditor, the auditor must take steps that minimize the likelihood of reaching an incorrect conclusion. So sampling minimizes the challenge of gathering sufficient and competent evidence for a given audit process. Furthermore, audit procedures are known to be expensive ventures to undertake at any given time, so examining all the population of transactions is merely impossible. Auditors routinely examine a representative sample of transactions on account balances. Failure to observe a representative sample may result in failure to detect material misstatement due to human error and the auditor is likely to reach a wrong conclusion about the population of transactions on the accounts based on the sample.

Since sampling is the basis of making a good inference about the population of transactions on the accounts, its accuracy depends on how representative the sample is of the population of transactions. Hence, when using samples that are not representative of the population of transaction the auditor must adjust the sample in a certain way in order to minimize the sampling risk inherent in the

audit assignment. Auditors can determine the sample size by examining the size of the material misstatement and by noting the difference to the auditor's overall assessment of correctness of the audit process. Auditors can also determine the sample size by using a confidence level based on the examination of sampled items.

The representativeness of a sample is used to assess the effectiveness of controls. If sampling is nonrepresentative, the auditor's risk of assessing controls may be too high, which leads to audit inefficiency, or assessing the control risk may be too low, resulting in ineffective control procedures hence increasing detection risk. There is also the risk of incorrect acceptance, which can lead to litigation, loss of reputation, and loss of clients [3].

1.2 Statistical Framework

Consider an account that contains a large population of transactions $X_1, \ldots, X_N$, each of which is a d dimension real-valued vector. Each transaction X_i contains d measurements that could contain information such as how much was charged, what type of procedure was performed, or other information. Each transaction also contains a component which equals the amount of fraud associated with the transaction. This last component will only be observed if that transaction receives an audit; otherwise this component is unobserved. In the context of sampling, suppose n transactions are sampled for auditing in the given population of size N.

Define X_{ij} to be the j^{th} component of the i^{th} transaction for $i = 1, \ldots, n$ and $j = 1, \ldots, d$. Let X_{i1} be the fraud associated with transaction i, where $i = 1, \ldots, n$. Therefore $(X_{i2}, \ldots, X_{id})'$ is the vector of covariates associated with transaction i, where $i = 1, \ldots, n$. As part of the problem, the covariates are known for all of the

transactions in the population, but the fraud amount is known only for a relatively small random sample of transactions through an audit. Let θ denote the amount of fraud in the population, which is given by

$$\theta = \sum_{i=1}^{N} X_{i1}.$$

Using standard techniques from sampling theory, the total fraud can be estimated by

$$\hat{\theta} = Nn^{-1} \sum_{i=1}^{n} X_{i1}.$$

This estimator is known to be an unbiased estimator with a standard error approximately equal to $\frac{N\sigma}{\sqrt{n}}$. The main question that we study is whether we get a better estimate of θ if we check our sample to insure that it is representative in some way. Let

$$\mu_j = N^{-1} \sum_{i=1}^{N} X_{ij}$$

for $j = 2, \ldots, d$, which is the true population mean for the j^{th} covariate. Let I_j denote a $100(1-\alpha)$ percent confidence interval for μ_j, based on, $X_{1j}, \ldots, X_{nj}$. This confidence interval is usually the standard *t-interval* for a population mean.

We wish to determine whether $\widehat{\theta}$ is closer to θ on average if we only accept a sample such that $\mu_j \in I_j$, for $j = 2, \ldots, d$. Other questions also arise that we will address, such as what value of α should be chosen, and whether $\widehat{\theta}$ can be adjusted

based on the covariates instead of using acceptance sampling.

1.3 Variance Reduction Techniques

Variance reduction techniques are experimental design and analysis techniques used to increase the precision of sampling-based point estimators without a corresponding increase in sampling effort [1]. Simulations driven by random inputs will always produce random outputs. Therefore it is imperative to apply appropriate statistical techniques to simulation output data for proper analysis and interpretation.

Variance reduction techniques were originally developed to be applied to Monte Carlo simulations or distribution sampling. Some examples of variance reduction techniques are stratified sampling, importance sampling, common random numbers, antithetic variates, control variates, induced estimation and conditioning. In this thesis we shall discuss the use of the control variate method of variance reduction as an alternative to the acceptance sampling plan for estimating the total fraud in financial audits. The key idea is that we know the mean of the covariates, and these means can be used to adjust the mean fraud estimate.

Correlations may arise naturally during the course of simulation or may be induced by the use of common random numbers in an auxiliary simulation. The method of control variates attempts to take advantage of correlation between certain random variables to obtain a reduction in the variance of an estimator of a mean. Suppose X is an output random variable and we want to estimate $\mu = E(X)$. Let Y be another random variable involved in the simulation that is correlated with X

(either positively or negatively), where we know the value of $V = E(Y)$. Momentarily suppose that the correlation between X and Y is positive. If we run a simulation and notice that $Y > V$, we might suspect that X is above its expectation μ as well, owing to the fact that we know the correlation between X and Y is positive, and accordingly adjust X downward by some amount. If we notice that $Y < V$, we would suspect $X < \mu$ as well and so adjust it upward. That is, we use the knowledge of $E(Y)$ to pull X (down or up) toward μ, thereby reducing its variability about μ in our simulation runs. So Y is our control variate for X since it is used to adjust X, or partially control it.

The success of control variate does not depend on the sign of the correlation. If X and Y were negatively correlated, we would simply adjust X upward if $Y > V$ and downward if $Y < V$. Hence monotonicity ensures that the correlations induced have the right sign for variance reduction.

To execute this routine it is convenient to express these trends in terms of the deviation $Y - V$, of Y from its expectation V. Define the controlled estimator to be

$$X_c = X - k(Y - V), \tag{1.1}$$

where k is a constant used to magnify or shrink the deviation $Y - V$ to arrive at an adjustment of X. Notice that if X is unbiased then X_c is also unbiased and might have lower variance than X. In particular

$$\mathrm{Var}(X_c) = \mathrm{Var}(X) + k^2\ \mathrm{Var}(Y) - 2k\ \mathrm{Cov}(X, Y). \tag{1.2}$$

So X_c is less variable than X if and only if

$$2k\ \mathrm{Cov}(X, Y) > k^2\ \mathrm{Var}(Y), \tag{1.3}$$

which may or may not be true depending on the choice of Y and k. By solving the previous equation for k we obtain the optimal variance-minimizing value given by

$$k^* = \frac{\text{Cov}(X, Y)}{\text{Var}(Y)}. \tag{1.4}$$

Using this optimal value of k, Equation (1.2) becomes

$$\text{Var}(X_c^*) = \text{Var}(X) - \frac{\text{Cov}^2(X, Y)}{\text{Var}(Y)} = (1 - \rho_{XY}^2)\text{Var}(X), \tag{1.5}$$

where ρ_{XY} is the correlation between X and Y.

We observe that by using the optimal value k^* for k, the optimally controlled estimator X_c^* cannot be more variable than the uncontrolled X. In particular X_c^* will have a lower variance than X if Y is correlated with X. Hence the stronger the correlation between X and Y, the greater the variance reduction.

In practice, we may not know the value of $\sigma^2(Y)$ and $\sigma(X, Y)$, making it impossible to find the exact value of k^*. [14], and [15] proposed a method of estimating k^*. Their method simply replaces $\sigma(X, Y)$ and $\sigma^2(Y)$ in Equation (1.4) by their sample estimators [5]. In this thesis we define X to be the *fraud amount* per transaction on each account and Y to be the *covariate* on the population of transaction on the accounts subject to audit for the acceptance sampling and control variate methodology. We assume there is a non-negative correlation between X and Y. We then apply the method proposed by [15].

The methodology of the study was detailed in Chapter 1. In Chapter 2, the simulation study is given for the acceptance sampling and control variate methodology.

The results of the simulation are discussed. A conclusion of the study and an area for future research is given in Chapter 3.

CHAPTER 2

SIMULATIONS

Multivariate statistics often involves the calculation of complex integrals whose analytic solutions cannot be easily evaluated in a closed form. In these cases we often resort to simulation to solve these problems. The most difficult integration occurs when the dimension d of the vector X becomes large. High dimensional integration is almost always difficult, even with the aid of simulations. In statistics, simulation methods are often called Monte Carlo methods, and high dimensional integration is often accomplished using Markov-chain Monte Carlo (MCMC) methods [9].

Almost all multivariate statistical methods assume some specific multivariate model for the underlying joint distribution. The most popular and the most often used multivariate model in practice is the multivariate normal distribution. This is due in part to its familiarity and convenience of simulating. Often observational data are non-normal, either skewed or relatively heavy-tailed. Unfortunately, the simulation of multivariate non-normal data is not a very common practice, partly because of the lack of algorithms for doing so [4].

In this thesis we use the statistical computing environment R for our simulations. The multivariate normal model was used as one model and a skewed normal mixture was used as another model. All simulations were performed, and all analysis, including graphics, was completed on a Microsoft Windows PC. Random variates were generated using *mvrnorm* package in the *MASS* library in R. The graphics were produced using the *KernSmooth* package.

The objective of our simulations is to estimate the standard error, bias, and mean squared error of the estimated mean fraud per transaction for the acceptance sampling and control variate methodology and compare them to the standard error of the standard methodology based on random sampling.

We define the following parameters. Let b be the number of replications on the account subject to audit, d be the dimension or the number of measurements taken on each account, ρ be the correlation between the components of each transaction, and n be the sample size drawn from each account.

We consider an account that contains a large population of transactions say in excess of ten thousand transactions. With this assumption the associated sampling without replacement is essentially independent. We consider dimensions equal to $d = 2$ and $d = 3$. Specific correlations are defined for the population of transactions. The values of correlations chosen are $\rho = 0.00$, $\rho = 0.25$, $\rho = 0.75$, and $\rho = 0.90$, and three sample sizes given by $n = 25$, $n = 50$, and $n = 100$, which can be considered as small, medium, and large samples, taken randomly from the population of transactions to be audited. Due to the assumption that the population of transactions is very large, we assume that no more than half of the accounts in any transaction may be audited [6]. That is, $n \leq b/2$. For the acceptance sampling algorithm, we use Student$-t$ confidence intervals for the covariate means with nominal levels $1 - \alpha = (0.25, 0.50, 0.75)$.

In the R code we set up the mean vector correlation matrix Σ by defining the following parameters:

```
m = rep(0,d)
I = diag (rep(1,d))
J = array (1, dim = c(d,d))
```

```
sigma = ((1 - rho) * I) + (rho * J).
```

Here ρ is taken to be the common correlation coefficient between the components of the covariates and the fraud. We set up a statistical counter to keep track of the number of samples that are accepted from simulation. The means from the accepted transactions are saved in a $1 \times d$ vector.

In the main simulation loop (a while loop), we generate a matrix X, using the *mvrnorm* package in the *MASS* library in *R*. The number of transactions (that is, the sample size) for each account, and the number of covariates are fixed. A *for loop* is used to check to see if the confidence interval for the mean of each covariate contains the *true mean* of the covariate for the transactions. If not, the sample is rejected, and another sample is taken. The objective here is to estimate the error amount in the population of transactions. Usually, the interest of the auditors is in obtaining a statistical upper bound for the true monetary error, which is likely to be greater than the error presented [7]. However, the estimated error may be less than the true amount desired. Whenever an estimated mean falls within the confidence interval it is saved and the statistical counter is increased by one. It is rejected otherwise. The simulation runs until $b = 10,000$ accepted means have been observed. Finally we compute the standard error for the usual estimate of the mean based on standard sampling assumptions, and the standard error for the validated samples. We compare these errors for all cases. We also compute the bias and mean squared error for each case.

Based on our simulation, the *standard errors* for the standard approach are 0.20 for $n = 25$, 0.14 for $n = 50$, and 0.10 for $n = 100$ for fixed values of ρ and α.
The simulation results are given in Tables $2.1 - 2.16$ below. We observe some differences in the results across the separate components of the validated error and

confidence intervals. We reported these components to see how they respond to changes in the sample size n of accounts subject to audit, measurement on each transaction d, the correlation between transactions ρ, and the confidence level α. When one focuses on the *validated error* column representing the error in the true mean transaction amount, one may notice the following. First, for $n = 25$, $d = 2$, and $\rho = 0.00, 0.25, 0.75$, and 0.90, the validated error increases from 0.0941 to about 0.6032. Such a result suggests that the validated error on the transactions account tends to increase when small samples are chosen and when the correlation ρ between the transactions is increased. The *validated bias* also increased from -0.0008 to 0.0029. The *validated RMSE* is also observed to have increased from about 0.0941 to about 0.2014. The largest *validated RMSE* value, 0.2014, is observed when $\rho = 0$ and $\alpha = 0.25$.

In a similar fashion, the proportion of means accepted in the acceptance sampling on the transaction on each account is observed to have increased from about 0.2495 to about 0.7549. This result strongly suggests that the likelihood of fraud detection in the audit accounts increases across components. Similarly, setting $n = 50$, $n = 100$, and $d = 2$ and increasing ρ while decreasing α, we observe a steady increase in the *validated error* as well as the *validated bias* and the *proportion of means accepted* in the acceptance sampling. These validated errors are higher when compared with the standard method based on random sampling.

The pattern of the effect of the correlation ρ on the validated errors varies across the components of α for given values of ρ, but are constant when the dimension of the covariate is increased. It is also worth noting that for each case considered, the validated errors are lower than the usual error from the standard method for the true monetary error. This result strongly suggests that the audit procedure based on this model may result in more effective litigation for the government agency.

In the *non-normal* case, Tables 2.9 to 2.14, the *validated bias* increased from -0.0259 to 0.0046 for $n = 25$, $d = 2$, and $\rho = 0.00, 0.25, 0.75$, and 0.90. By setting $n = 25$ and $d = 2$ and increasing ρ while decreasing α, we observe a steady decrease in the *validated error*, whereas the *validated bias* and the *proportion of means accepted* in the acceptance sampling are increasing. These validated errors are higher when compared with the standard method based on random sampling. A similar trend is observed when the sample size $n = 50$, and $n = 100$. This result suggests that the fraud detection in the audit account decreases across components. Hence the audit procedure based on the *non-normal* model may result in less effective litigation for the government agency.

We observe from the tables of control variates, Tables 2.7 to 2.8, that the non-adjusted root mean square error (non-adjusted $RMSE$) decreases from 0.20 to 0.10 as the sample size increases from $n = 25$ to $n = 100$ for fixed values of correlation ρ $(= 0, 0.25, 0.75, 0.90)$. The adjusted root mean square error (adjusted $RMSE$) decreases from 0.2037 to 0.0435 when the dimension of measurement on the accounts $d = 2$. Hence increasing the dimension of measurements on the population of accounts to be audited reduces the *adjusted RMSE* on the accounts whenever the sample taken on each account is increased and provided the correlation on the accounts are fixed at ρ $(= 0, 0.25, 0.75, 0.90)$. We also observe from the tables of control variates that the adjusted bias on each account is very small (or essentially zero). We observe that the adjusted root mean squares error *(RMSE)* decreases from 0.2037 to 0.0435 when the measurement on each account $d = 2$ and the *RMSE* decreases from 0.2049 to 0.0437 when $d = 3$. The *adjusted RMSE* was observed to be lower than the *non-adjusted RMSE* in almost all the data sets at all sample sizes. This suggests that the audit process based on the *normal* model may result in more

effective litigation for the government agency. In each case the control variate errors compares lower to the acceptance sampling errors.

Adjusted RMSE comparisons were consistent across specific correlations ρ, with *adjusted RMSE* for $\rho = 0$ always greater than *adjusted RMSE* for $\rho = 0.25$. Similar trend was observed as ρ increases irrespective of the sample size taken on each account transaction. In some cases, the *adjusted RMSE* was significantly larger than the *non-adjusted RMSE*.

In the *non-normal* case, Tables 2.15 to 2.16, we observe that the adjusted root mean square error (adjusted $RMSE$) decreases from 0.5266 to 0.0439 when the dimension of measurement on the accounts $d = 2$ and the $RMSE$ decreases from 0.2049 to 0.0437 when $d = 3$. The *adjusted RMSE* was observed to be higher than the *non-adjusted RMSE* in most of the data sets at fixed sample sizes. This suggests that the audit process based on the *non-normal* model may result in less effective litigation for the government agency.

Table 2.1: Simulation Results, Normal Acceptance Sampling for $n = 25$, $d = 2$.

ρ	$1-\alpha$	Val. Error	Val. Bias	Val. RMSE	Proportion
0.00	0.25	0.2014	0.0006	0.2014	0.2495
	0.50	0.2008	-0.0015	0.2008	0.4987
	0.75	0.6032	-0.0001	0.1990	0.7493
0.25	0.25	0.1965	-0.0006	0.1931	0.254
	0.50	0.1951	-0.0002	0.1951	0.4982
	0.75	0.1968	0.0014	0.1968	0.7453
0.75	0.25	0.1347	0.0029	0.1348	0.2521
	0.50	0.1435	0.0012	0.1435	0.4965
	0.75	0.1618	-0.0008	0.1618	0.7517
0.90	0.25	0.0941	0.0007	0.0941	0.2499
	0.50	0.1124	-0.0002	0.1124	0.5051
	0.75	0.5964	0.0009	0.1398	0.7548

Table 2.2: Simulation Results, Normal Acceptance Sampling for $n = 50$, $d = 2$.

ρ	$1-\alpha$	Val. Error	Val. Bias	Val. RMSE	Proportion
0.00	0.25	0.1429	0.0023	0.1430	0.2496
	0.50	0.1424	-0.0008	0.1424	0.5004
	0.75	0.1416	-0.0008	0.1416	0.7537
0.25	0.25	0.1367	-0.0023	0.1368	0.2532
	0.50	0.1379	0.0014	0.1379	0.5046
	0.75	0.1382	-0.0008	0.1382	0.7498
0.75	0.25	0.0955	0.0001	0.0955	0.2518
	0.50	0.1027	0.0008	0.1027	0.5044
	0.75	0.1152	-0.0003	0.1152	0.7488
0.90	0.25	0.0666	-0.0001	0.0666	0.2511
	0.50	0.0783	0.0004	0.0783	0.5002
	0.75	0.1009	-0.0004	0.1009	0.7468

Table 2.3: Simulation Results, Normal Acceptance Sampling for $n = 100$, $d = 2$.

ρ	$1-\alpha$	Val. Error	Val. Bias	Val. RMSE	Proportion
0.00	0.25	0.1003	0.0012	0.1003	0.2492
	0.50	0.1002	0.0000	0.1002	0.4980
	0.75	0.1005	0.0017	0.1006	0.7506
0.25	0.25	0.0968	0.0025	0.0969	0.2495
	0.50	0.0971	0.0010	0.0971	0.4982
	0.75	0.0977	-0.0014	0.0977	0.7483
0.75	0.25	0.0723	0.0012	0.0681	0.2478
	0.50	0.2917	0.0001	0.07228	0.4970
	0.75	0.0804	0.0002	0.0804	0.74702
0.90	0.25	0.0465	-0.0006	0.0465	0.2474
	0.50	0.0559	-0.0008	0.0559	0.5053
	0.75	0.0703	0.0002	0.0703	0.7521

Table 2.4: Simulation Results, Normal Acceptance Sampling for $n = 25$, $d = 3$.

ρ	$1-\alpha$	Val. Error	Val. Bias	Val. RMSE	Proportion
0.00	0.25	0.2008	0.0031	0.2008	0.0622
	0.50	0.1993	0.0006	0.1993	0.2520
	0.75	0.2009	-0.0003	0.2009	0.5676
0.25	0.25	0.1876	-0.0008	0.1876	0.0642
	0.50	0.1902	-0.0026	0.1903	0.2586
	0.75	0.1500	0.0010	0.1920	0.5652
0.75	0.25	0.1220	-0.0018	0.1220	0.0909
	0.50	0.1296	0.0002	0.1296	0.3241
	0.75	0.1470	-0.0016	0.1470	0.6358
0.90	0.25	0.0823	-0.0004	0.0823	0.1269
	0.50	0.0964	0.0001	0.0964	0.3864
	0.75	0.1287	-0.0015	0.1287	0.6724

Table 2.5: Simulation Results, Normal Acceptance Sampling for $n = 50$, $d = 3$.

ρ	$1-\alpha$	Val. Error	Val. Bias	Val. RMSE	Proportion
0.00	0.25	0.1404	0.0001	0.1404	0.0626
	0.50	0.1399	0.0015	0.1399	0.2495
	0.75	0.1426	-0.0010	0.1427	0.5615
0.25	0.25	0.1350	0.0011	0.1350	0.0648
	0.50	0.1347	0.0006	0.1347	0.2555
	0.75	0.1360	-0.0007	0.1360	0.5681
0.75	0.25	0.0863	-0.0012	0.0863	0.0863
	0.50	0.0913	0.0001	0.0913	0.3241
	0.75	0.1044	-0.00001	0.1044	0.6380
0.90	0.25	0.0581	-0.0007	0.0581	0.1269
	0.50	0.0681	0.0010	0.0681	0.3876
	0.75	0.0910	-0.0001	0.0910	0.6789

Table 2.6: Simulation Results, Normal Acceptance Sampling for $n = 100$, $d = 3$.

ρ	$1-\alpha$	Val. Error	Val. Bias	Val. RMSE	Proportion
0.00	0.25	0.0995	0.0007	0.0995	0.0627
	0.50	0.1008	-0.0011	0.1008	0.2533
	0.75	0.1008	-0.0013	0.1008	0.5592
0.25	0.25	0.0961	-0.0003	0.0961	0.0653
	0.50	0.0960	0.0003	0.0960	0.2580
	0.75	0.0961	0.0005	0.0961	0.5747
0.75	0.25	0.0611	-0.0006	0.0611	0.0932
	0.50	0.0651	-0.0002	0.0651	0.3265
	0.75	0.0737	-0.0007	0.0737	0.6321
0.90	0.25	0.0405	0.0005	0.0405	0.1249
	0.50	0.0483	0.0009	0.0484	0.3809
	0.75	0.0638	0.0001	0.0639	0.6752

Table 2.7: Control Variate Results for Normal Model, $d = 2$.

n	d	ρ	$RMSE$	$Adjusted - SE$	$Adjusted - Bias$	$Adjusted - RMSE$
25	2	0.00	0.20	0.2037	0.0001	0.2037
		0.25	0.20	0.1969	-0.0045	0.1969
		0.75	0.20	0.1342	0.0004	0.1342
		0.90	0.20	0.0892	-0.0001	0.0892
50	2	0.00	0.14	0.1443	-0.0011	0.1443
		0.25	0.14	0.1384	-0.0014	0.1384
		0.75	0.14	0.1253	-0.0025	0.1253
		0.90	0.14	0.0623	0.0006	0.0623
100	2	0.00	0.10	0.1008	0.00165	0.1008
		0.25	0.10	0.0964	-0.0009	0.0964
		0.75	0.10	0.0657	0.0001	0.0657
		0.90	0.10	0.0435	-0.0006	0.0435

Table 2.8: Control Variate Results for Normal Model, $d = 3$.

n	d	ρ	$RMSE$	$Adjusted - SE$	$Adjusted - Bias$	$Adjusted - RMSE$
25	3	0.00	0.20	0.2033	-0.0009	0.2033
		0.25	0.20	0.1989	0.0020	0.1989
		0.75	0.20	0.1360	-0.0019	0.1360
		0.90	0.20	0.1360	-0.0019	0.1360
50	3	0.00	0.14	0.1424	0.0017	0.1424
		0.25	0.14	0.1375	0.0016	0.1376
		0.75	0.14	0.0945	-0.0010	0.0945
		0.90	0.14	0.0621	-0.0012	0.0621
100	3	0.00	0.10	0.0998	0.0003	0.0998
		0.25	0.10	0.0982	0.0018	0.0982
		0.75	0.10	0.0662	0.0004	0.0662
		0.90	0.10	0.0439	0.0003	0.0439

Table 2.9: Simulation Results, Non-Normal Acceptance Sampling for $n = 25$, $d = 2$.

ρ	α	Val. Error	Val. Bias	Val. RMSE	Proportion
0.00	0.25	0.6048	-0.0043	0.6048	0.9795
	0.50	0.5208	-0.0017	0.5208	0.5017
	0.75	0.5408	-0.0184	0.5411	0.7549
0.25	0.25	0.5121	-0.0022	0.5121	0.2514
	0.50	0.5244	-0.0172	0.5247	0.4989
	0.75	0.5499	-0.027	0.5505	0.7484
0.75	0.25	0.5113	0.0046	0.5114	0.2455
	0.50	0.5213	-0.0058	0.5214	0.4947
	0.75	0.5431	-0.0259	0.5437	0.7458
0.90	0.25	0.5082	0.0031	0.5082	0.2472
	0.50	0.5223	-0.0158	0.5226	0.4972
	0.75	0.5482	-0.0204	0.5486	0.7454

Table 2.10: Simulation Results, Non-Normal Acceptance Sampling for $n = 50$, $d = 2$.

ρ	$1 - \alpha$	Val. Error	Val. Bias	Val. RMSE	Proportion
0.00	0.25	0.3613	0.0032	0.3614	0.2505
	0.50	0.3718	0.0014	0.3718	0.5037
	0.75	0.3834	–0.0142	0.3837	0.7498
0.25	0.25	0.3585	0.0015	0.3585	0.2538
	0.50	0.3681	-0.0067	0.3682	0.4979
	0.75	0.3838	-0.0053	0.3838	0.7479
0.75	0.25	0.3591	-0.0041	0.3591	0.2516
	0.50	0.3700	-0.0004	0.3699	0.4996
	0.75	0.3868	-0.0080	0.3868	0.7551
0.90	0.25	0.3588	-0.0045	0.3589	0.2505
	0.50	0.3667	-0.0064	0.3668	0.4979
	0.75	0.3827	-0.0086	0.3828	0.7507

Table 2.11: Simulation Results, Non-Normal Acceptance Sampling for $n = 100$, $d = 2$.

ρ	$1-\alpha$	Val. Error	Val. Bias	Val. RMSE	Proportion
0.00	0.25	0.2572	0.0048	0.2572	0.2530
	0.50	0.0995	-0.0010	0.0995	0.5024
	0.75	0.2619	-0.0035	0.2619	0.5067
0.25	0.25	0.2539	0.0029	0.2539	0.2472
	0.50	0.2596	-0.0021	0.2597	0.5019
	0.75	0.2720	-0.0038	0.2720	0.7502
0.75	0.25	0.2557	-0.0024	0.2557	0.2513
	0.50	0.2629	-0.0031	0.2629	0.5017
	0.75	0.2732	-0.0076	0.2733	0.7520
0.90	0.25	0.2572	0.0054	0.2572	0.2512
	0.50	0.2639	0.0021	0.2639	0.4979
	0.75	0.2732	-0.0062	0.2733	0.7482

Table 2.12: Simulation Results, Non-Normal Acceptance Sampling for $n = 25$, $d = 3$.

ρ	$1-\alpha$	Val. Error	Val. Bias	Val. RMSE	Proportion
0.00	0.25	0.4969	-0.0141	0.4971	0.0630
	0.50	0.5076	-0.0154	0.5078	0.2529
	0.75	0.5435	-0.0230	0.5439	0.5666
0.25	0.25	0.4976	-0.0045	0.4976	0.0644
	0.50	0.5146	-0.0073	0.5147	0.2546
	0.75	0.5417	-0.0262	0.5423	0.5667
0.75	0.25	0.4962	-0.0068	0.4962	0.0639
	0.50	0.5137	-0.0124	0.5139	0.2547
	0.75	0.5425	-0.0108	0.5426	0.5586
0.90	0.25	0.4994	-0.0049	0.4994	0.0628
	0.50	0.5089	-0.0222	0.5094	0.2556
	0.75	0.5442	-0.0258	0.5448	0.5612

Table 2.13: Simulation Results, Non-Normal Acceptance Sampling for $n = 50$, $d = 3$.

ρ	$1-\alpha$	Val. Error	Val. Bias	Val. RMSE	Proportion
0.00	0.25	0.3462	0.0021	0.3462	0.0626
	0.50	0.3585	-0.0144	0.3588	0.2531
	0.75	0.3813	-0.0079	0.3814	0.5696
0.25	0.25	0.3452	-0.0097	0.3453	0.0640
	0.50	0.3623	-0.0028	0.3623	0.2493
	0.75	0.3779	-0.0106	0.3780	0.5683
0.75	0.25	0.3478	-0.0072	0.3479	0.0630
	0.50	0.3595	0.0015	0.3595	0.2511
	0.75	0.3829	-0.0098	0.3830	0.5595
0.90	0.25	0.3507	-0.0063	0.3508	0.0627
	0.50	0.3646	0.0022	0.3646	0.2544
	0.75	0.3821	-0.0087	0.3822	0.5626

Table 2.14: Simulation Results, Non-Normal Acceptance Sampling for $n = 100$, $d = 3$.

ρ	$1-\alpha$	Val. Error	Val. Bias	Val. RMSE	Proportion
0.00	0.25	0.2461	-0.0032	0.2461	0.0632
	0.50	0.2582	-0.0011	0.2582	0.2534
	0.75	0.2699	–0.0016	0.2699	0.5695
0.25	0.25	0.2446	-0.0009	0.2446	0.0617
	0.50	0.2536	-0.0041	0.2537	0.2538
	0.75	0.2699	-0.0049	0.2700	0.5647
0.75	0.25	0.2499	-0.0024	0.2499	0.0634
	0.50	0.2552	-0.0030	0.2552	0.2494
	0.75	0.3829	-0.0098	0.3830	0.5595
0.90	0.25	0.2461	-0.0010	0.2461	0.0619
	0.50	0.2564	-0.0043	0.2565	0.2567
	0.75	0.2722	-0.0044	0.2722	0.5600

Table 2.15: Control Variate Results for Non-Normal Model, $d = 2$.

n	d	ρ	$RMSE$	$Adjusted - SE$	$Adjusted - Bias$	$Adjusted - RMSE$
25	2	0.00	0.20	0.5194	-0.0415	0.5210
		0.25	0.20	0.5183	-0.0365	0.5195
		0.75	0.20	0.5178	-0.0415	0.5195
		0.90	0.20	0.5215	-0.0343	0.5226
50	2	0.00	0.14	0.3598	-0.0156	0.3601
		0.25	0.14	0.3602	-0.0148	0.3606
		0.75	0.14	0.3641	-0.0232	0.3649
		0.90	0.14	0.3581	-0.0112	0.3582
100	2	0.00	0.10	0.0998	0.0003	0.0998
		0.25	0.10	0.0982	0.0018	0.0982
		0.75	0.10	0.0662	0.0004	0.0662
		0.90	0.10	0.0439	0.0003	0.0439

Table 2.16: Control Variate Results for Non-Normal Model, $d = 3$.

n	d	ρ	$RMSE$	$Adjusted - SE$	$Adjusted - Bias$	$Adjusted - RMSE$
25	3	0.00	0.20	0.2048	-0.0029	0.2049
		0.25	0.20	0.1965	0.0003	0.1965
		0.75	0.20	0.1326	0.0002	0.1326
		0.90	0.20	0.0892	-0.0004	0.0892
50	3	0.00	0.14	0.1420	0.0007	0.1420
		0.25	0.14	0.1393	0.0006	0.1393
		0.75	0.14	0.0948	0.0010	0.0948
		0.90	0.14	0.0620	0.0005	0.0620
100	3	0.00	0.10	0.1002	0.0006	0.1002
		0.25	0.10	0.0973	-0.0007	0.0973
		0.75	0.10	0.0666	0.0008	0.0666
		0.90	0.10	0.0437	-0.0007	0.0437

The graphical illustrations presented exemplify the equicorrelation and exchangeability of the components of the simulated multivariate normal model through histograms. The random numbers needed are generated by using R function $mvrnorm$. Figure 2.1 through Figure 2.12 contains twelve pairwise histograms. One is the histogram of the unaccepted means and the other is for the accepted means for simulated samples of size $n = 25$, $n = 50$, and $n = 100$, respectively, from a multivariate normal distribution.

Figure 2.1 through Figure 2.7 show unimodal and symmetric histograms centered at zero. This is an indication of little or no fraud in our acceptance sampling. Figure 2.8 through Figure 2.12 show two distinct characteristics. Observe that the histograms for the unaccepted means are bimodal. The bimodality of the unaccepted means becomes more distinct as the dimension of measurement d, and sample size n increases (see details in Appendix A). We observe that the center of these histograms for the unaccepted means are different from zero. This is an indication of the presence of some fraud in our acceptance sampling.

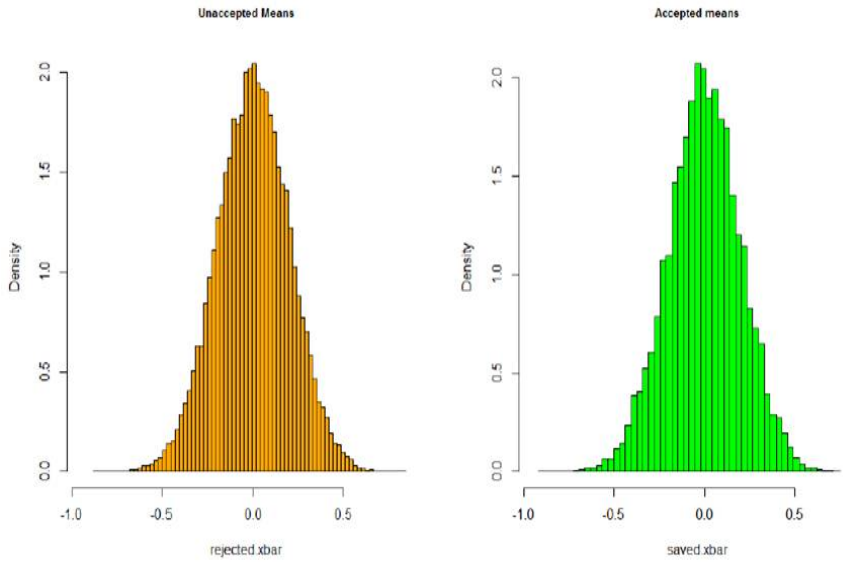

Figure 2.1: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0$, $\alpha = 0.25$.

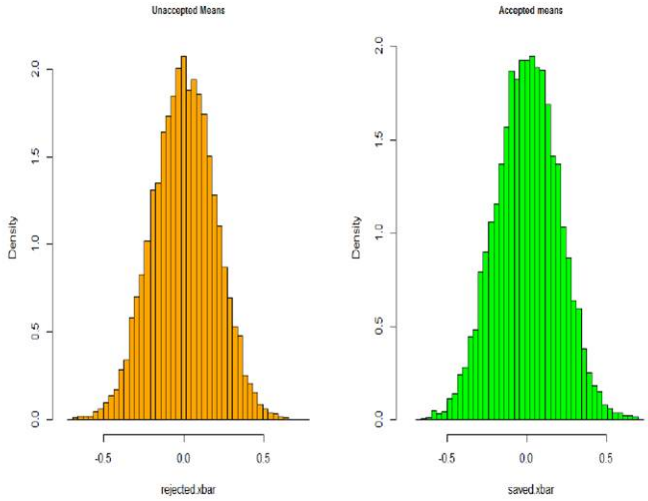

Figure 2.2: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0$, $\alpha = 0.50$.

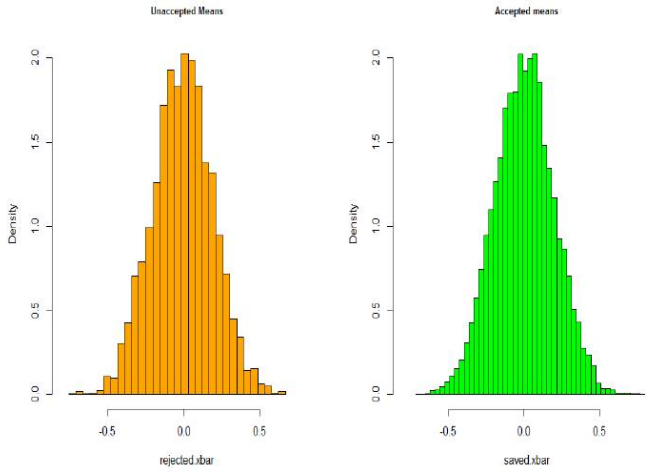

Figure 2.3: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0$, $\alpha = 0.75$.

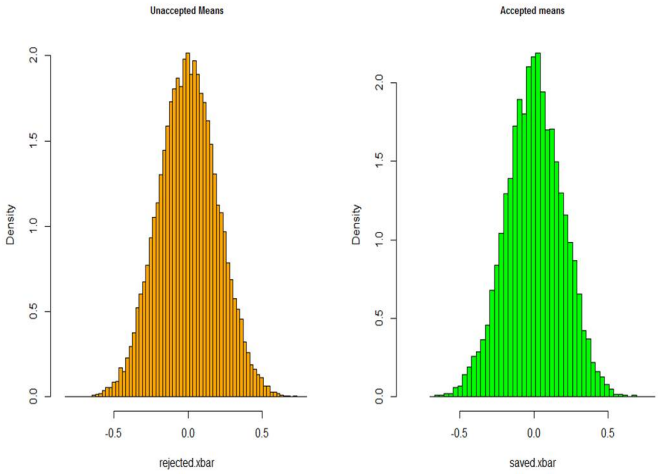

Figure 2.4: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.50$, $\alpha = 0.25$.

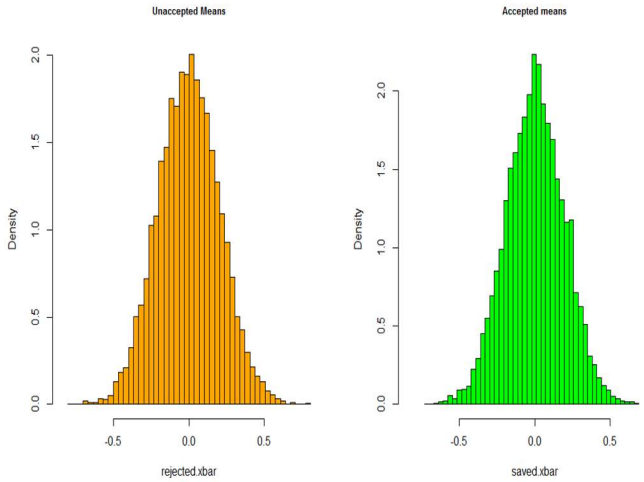

Figure 2.5: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.50$, $\alpha = 0.50$.

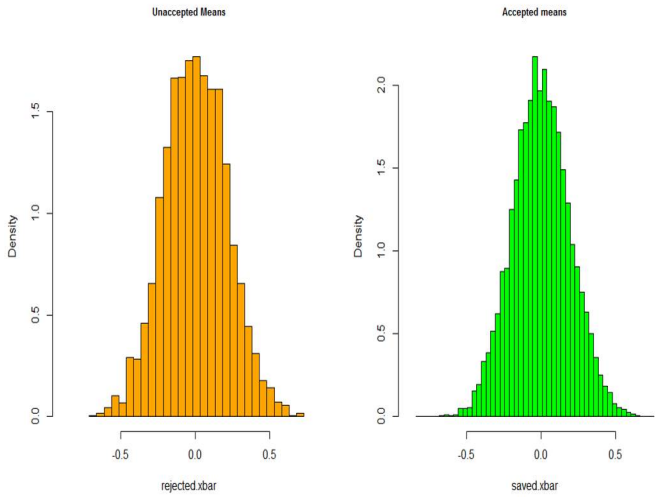

Figure 2.6: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.50$, $\alpha = 0.75$.

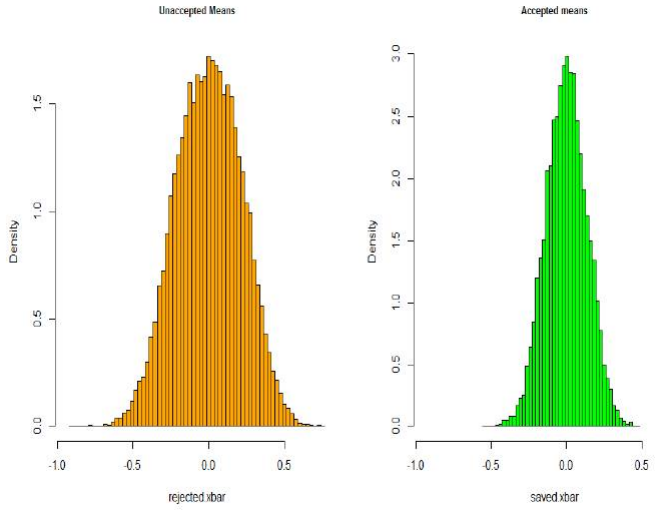

Figure 2.7: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.75$, $\alpha = 0.25$.

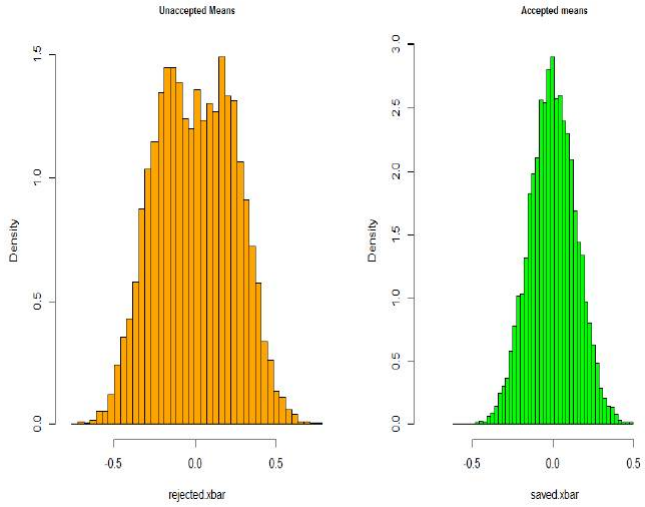

Figure 2.8: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.75$, $\alpha = 0.50$.

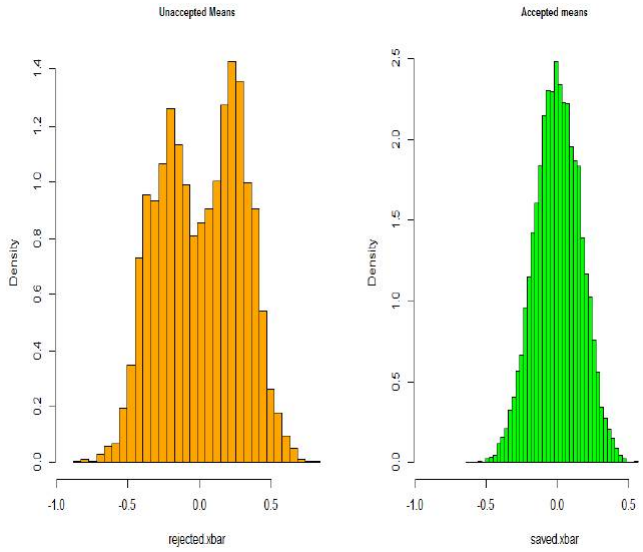

Figure 2.9: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.75$, $\alpha = 0.75$.

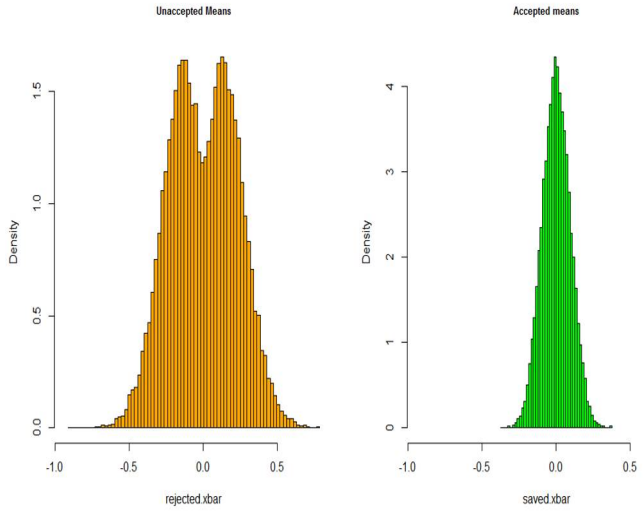

Figure 2.10: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.90$, $\alpha = 0.25$.

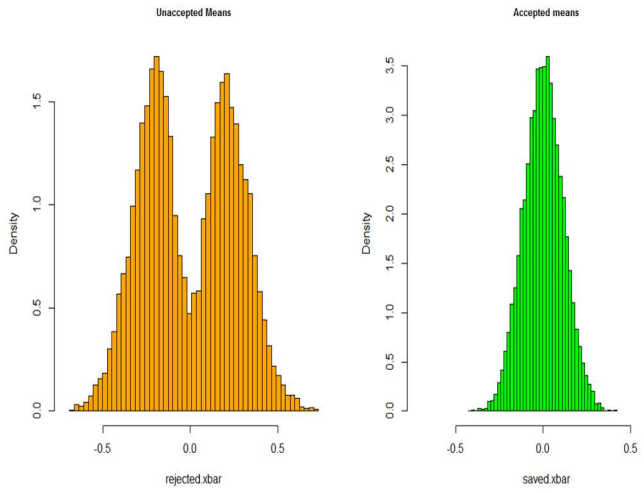

Figure 2.11: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.90$, $\alpha = 0.50$.

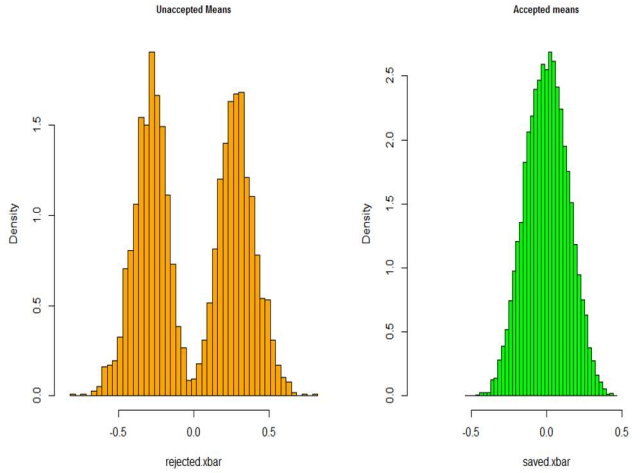

Figure 2.12: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.90$, $\alpha = 0.75$.

CHAPTER 3

CONCLUSION

The purpose of this thesis was to use computer-based simulations to study the effect of fraud on accounts transactions of a state government agency using a representative sample and illustrate how the control variate method can be used to reduce the estimation error for multivariate normal distribution. The *validated error*, *validated bias*, validated root mean square error *(validated RMSE)*, and *proportion of means accepted* in the acceptance sampling were determined. The adjusted standard error *(adjusted SE)*, *adjusted bias*, and the *adjusted RMSE* were also determined.

From our numerical experiments it was observed that the estimation error for the multivariate normal distribution is lower using our proposed method when compared with standard methods.

In summary, we observe the following:

- Validation reduces the error over standard sampling when the population is normal.
- Validation does not work as well when the population is not normal.
- Control variate method reduces error over standard sampling for both the *normal* and *non-normal* models.
- Control variate errors compares lower to the validation errors for all cases.

CHAPTER 4

FUTURE RESEARCH

Additional research could be explored in this area to:

- Derive theoretical results of the estimation error for both *normal* and *non-normal* distributions.

- Incorporate nonstandard mixture distributions. Nonstandard mixtures are those that result from a mixture of a discrete and a continuous random variable [11]. In particular, note that in auditing, some population accounts contain no errors (that is, the fraud is zero), whereas other accounts contain errors, where the fraud follows a continuous distribution [7]. Suppose E and F are the distribution of errors and no errors, respectively. Let $0 \leq X \leq 1$ and define $X_1 = XE + (1 - X)F$, to be a mixing distribution of E and F. Thus with probability X, we observe X_1 having distribution E, and with probability $1 - X$, we observe X_1 having distribution F.

- Extend our research to higher dimension of measurements (say $d \geq 5$) on the population of accounts subject to audit for both multivariate *normal* and *non-normal* distributions.

- Consider different types of confidence intervals (the $t-interval$ is not the only confidence interval for the mean).

REFERENCES

[1] Nelson, B. L. (1989). *Control Variate Remedies*. Ohio State University, Columbus, Ohio.

[2] Burke, K. D. (2010*a*). *Audit Sampling*. Available at `http://home.millsaps.edu/burkekg/AUDCH10.htm`.

[3] Burke, K. D. (2010*b*). *Audit Sampling*. Available at `http://home.millsaps.edu/burkekg/AUDCH9.htm`.

[4] Minhajuddin, A.T. M., Harris, I.R., and Schucany, W.R. (2003). *Simulating Multivariate Distributions with Specific Correlations*. Department of Statistical Science, Southern Methodist University, Dallas, Texas.

[5] Law, A.M., and Kelton, W.D. (2000). *Simulation Modeling and Analysis*, 3rd edition. McGraw-Hill.

[6] Heiner, K.W., Kupferschmid, M., Ecker, J., and Jordan, J.G. (1983). Maximizing Restitution for Erroneous Medical Payments When Auditing Samples from More Than One Provider. *J. Informs*, Vol. 13, No. 5, pp. $12 - 17$.

[7] Statistical Models and Analysis in Auditing: Panel on Nonstandard Mixtures of Distributions. (1989). *Statistical Science*, Vol. 4, No. 1, pp. $2 - 33$.

[8] State Medicaid Fraud Control Units Fiscal Year (2011). Grant Expenditures and Statistics. Available at `https://oig.hhs.gov/fraud/medicaid-fraud-control-units-mfcu/expenditures_statistics/fy2011.asp`.

[9] Christensen, R., Johnson, W., Branscum, A., and Hanson, T. E. (2011). *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians*. CRC Press Boca Raton, FL.

[10] United States Government Accountability Office (2006). *Financial Auditing. Significant Internal Control Weaknesses Remain in Preparing the Consolidated Financial Statements of the U.S. Government GAO*-06 − 415. Available at `http://www.gao.gov/products/GAO-`06 − 415.

[11] Polansky, A. M. (2005). Nonparametric Estimation of Distribution Functions of Nonstandard Mixtures. *Communications in Statistics-Theory and Methods*, vol. 34, no. 8, pp. 1711 − 1724.

[12] What Is Enron? *Clear Answers for Common Questions*. Available at `http://www.wisegeek.com/what-is-enron.htm`.

[13] Lucci, J. P. (2003). *EnronThe Bankruptcy Heard Around The World and The International Ricochet of Sarbanes-Oxley*. Available at `http://www.albanylawreview.org/archives/67/1/ENRON-THE BANKRUPTCYHEARDAROUNDTHEWORLDANDTHEINTERNATIONALRICOCHETOFSARBANE S-OXLEY.pdf, pp.` 211.

[14] Lavenberg, S.S., Moeller, T.L., and Welch, P.D. (1982).

Statistical results on multiple control variables with application to queueing network simulation. *Oper. Res.*, vol. 30, pp. 182 − 202.

[15] Lavenberg, S.S., and Welch, P.D. (1981). A perspective on the use of control variables to increase the efficiency of Monte Carlo simulations. *Manage. Sci.*, vol. 27, pp. 322 − 335

APPENDIX A

HISTOGRAM PLOTS

Plots of all histograms of unaccepted means and accepted means from our simulation.

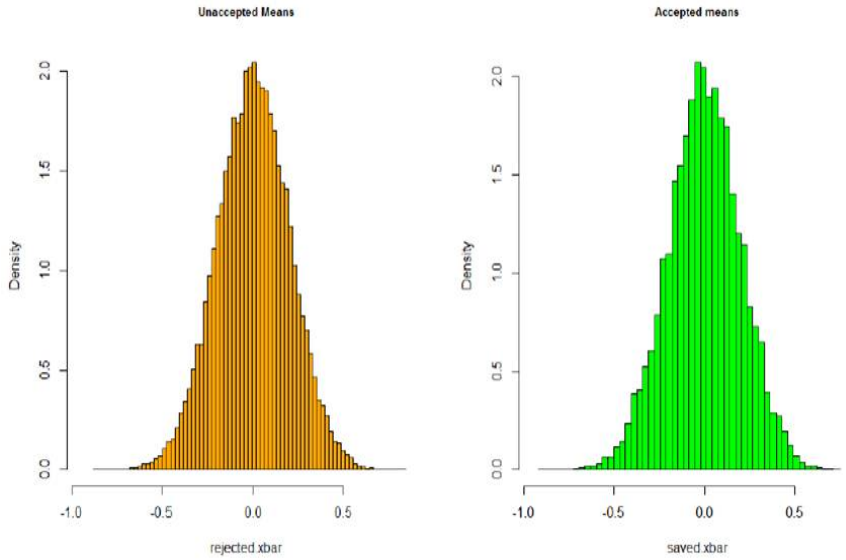

Figure A.1: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0$, $\alpha = 0.25$.

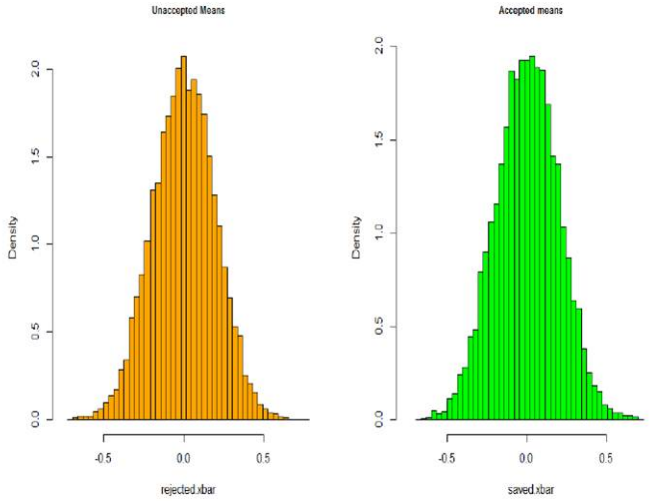

Figure A.2: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0$, $\alpha = 0.50$.

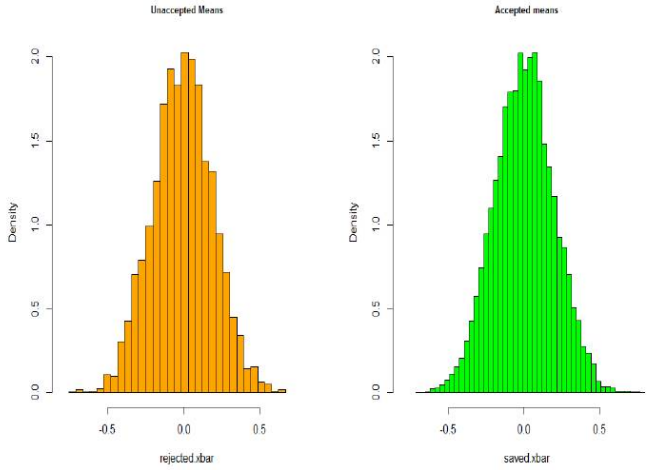

Figure A.3: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0$, $\alpha = 0.75$.

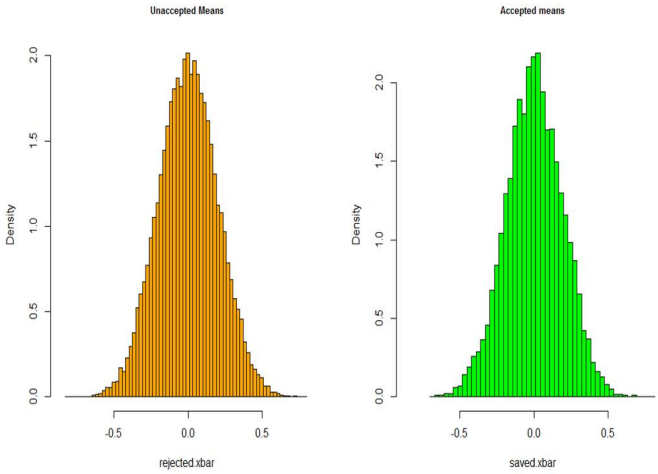

Figure A.4: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.50$, $\alpha = 0.25$.

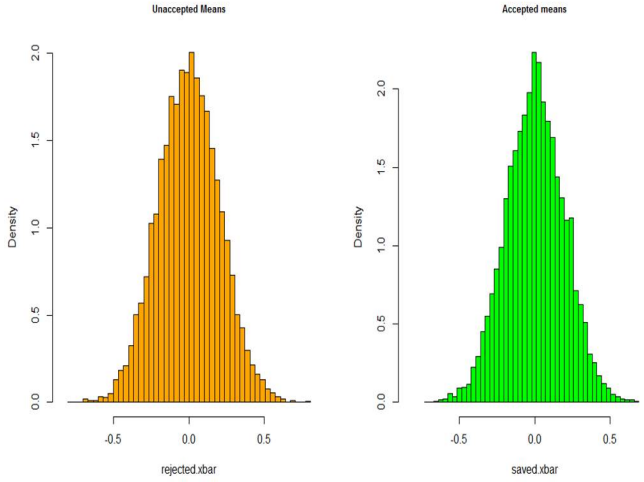

Figure A.5: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.50$, $\alpha = 0.50$.

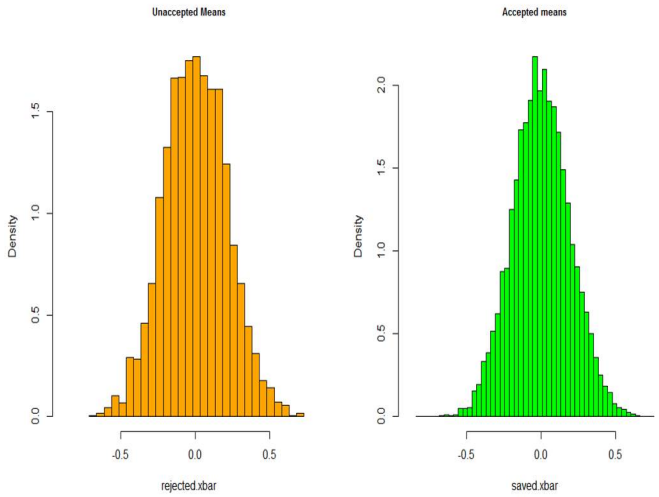

Figure A.6: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.50$, $\alpha = 0.75$.

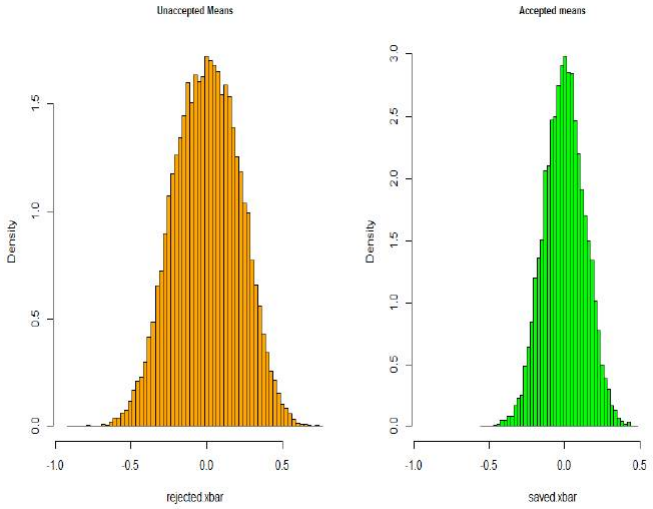

Figure A.7: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.75$, $\alpha = 0.25$.

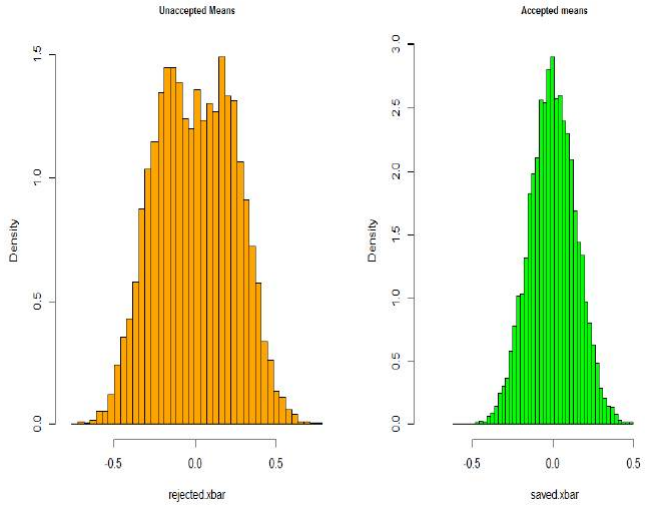

Figure A.8: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.75$, $\alpha = 0.50$.

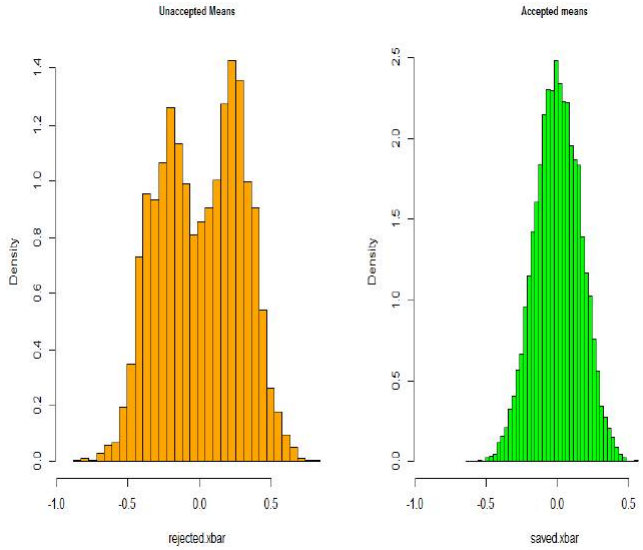

Figure A.9: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.75$, $\alpha = 0.75$.

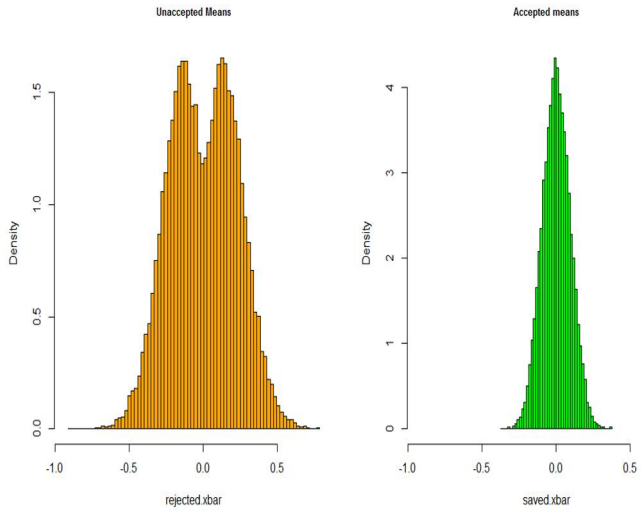

Figure A.10: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.90$, $\alpha = 0.25$.

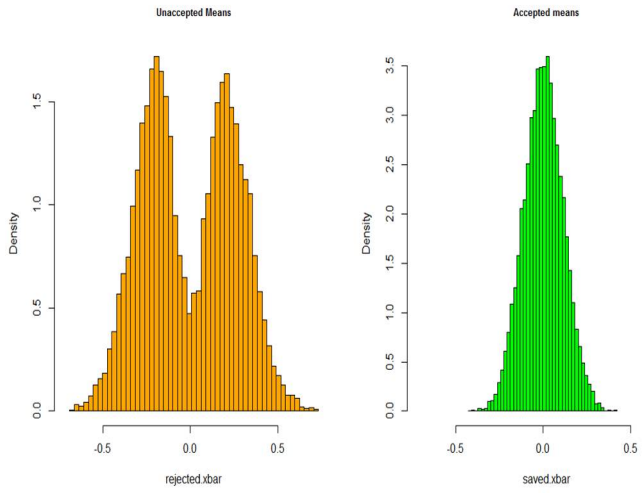

Figure A.11: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.90$, $\alpha = 0.50$.

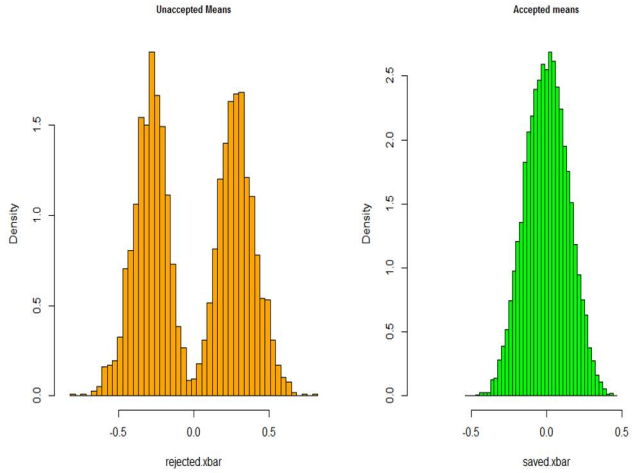

Figure A.12: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 2$, $\rho = 0.90$, $\alpha = 0.75$.

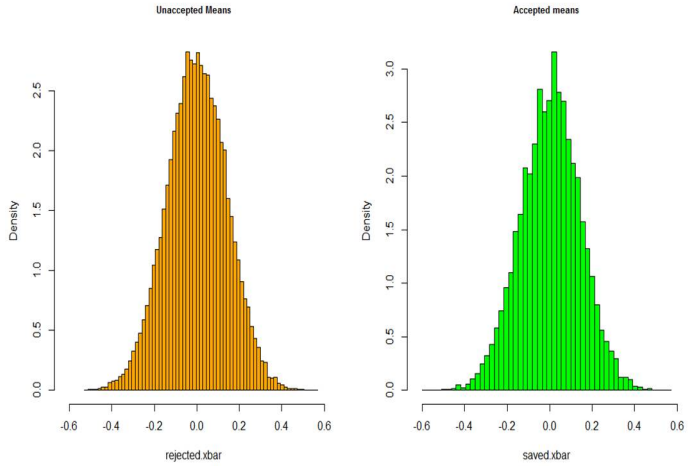

Figure A.13: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 2$, $\rho = 0$, $\alpha = 0.25$.

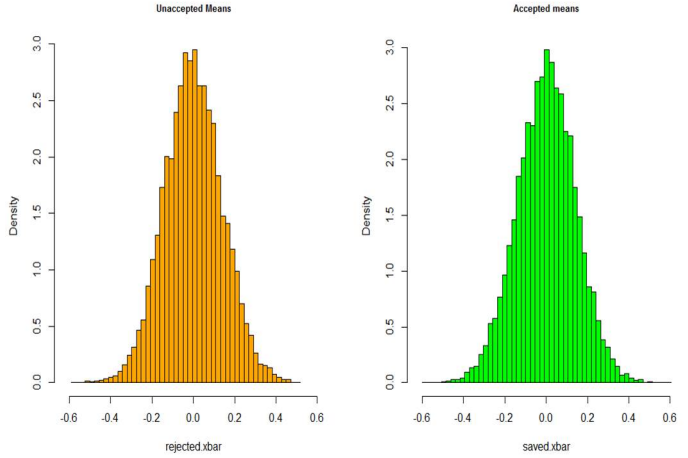

Figure A.14: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 2$, $\rho = 0$, $\alpha = 0.50$.

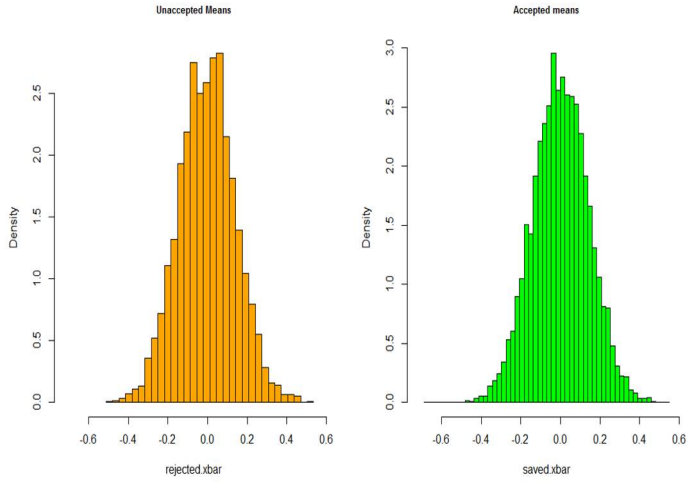

Figure A.15: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 2$, $\rho = 0$, $\alpha = 0.75$.

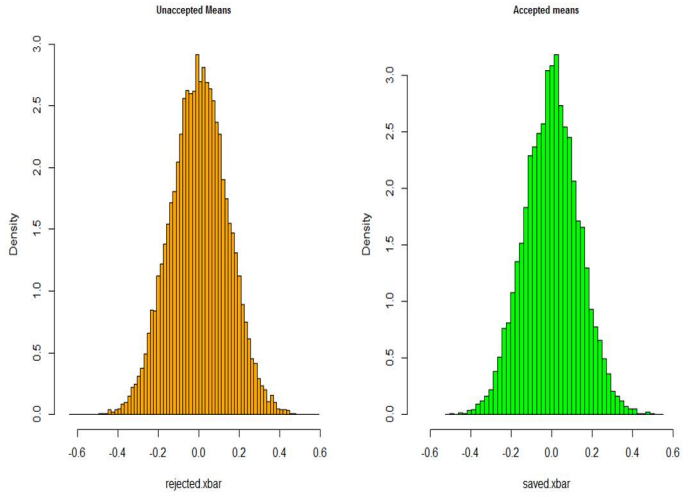

Figure A.16: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 2$, $\rho = 0.25$, $\alpha = 0.25$.

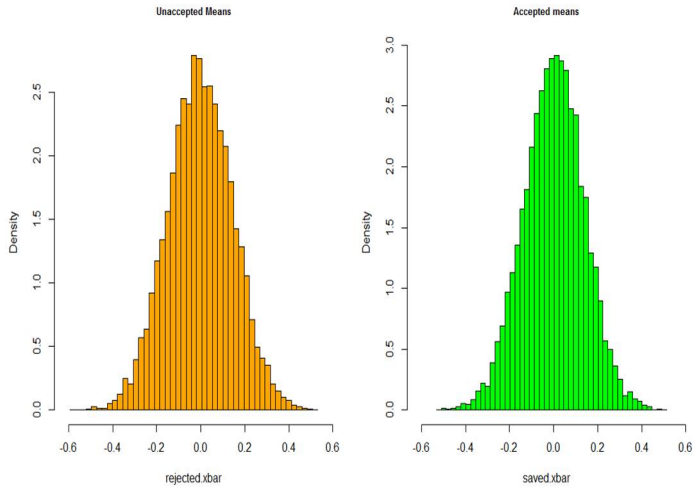

Figure A.17: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 2$, $\rho = 0.25$, $\alpha = 0.50$.

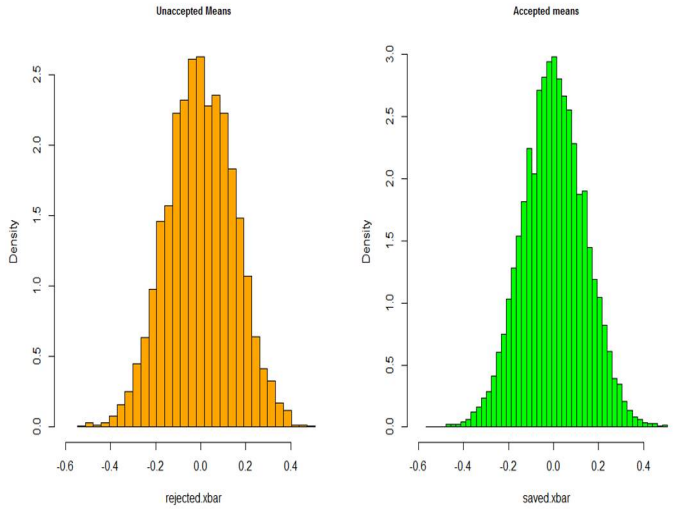

Figure A.18: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 2$, $\rho = 0.25$, $\alpha = 0.75$.

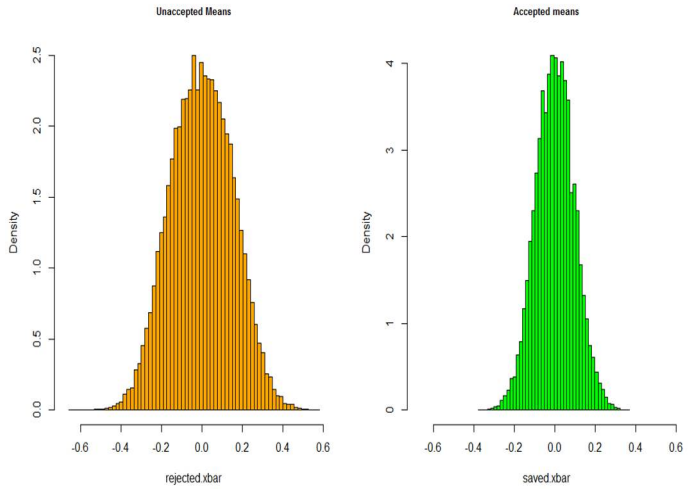

Figure A.19: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 2$, $\rho = 0.75$, $\alpha = 0.25$.

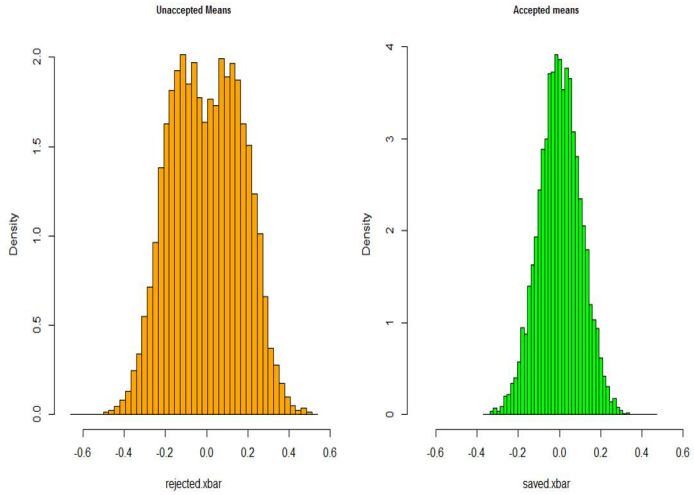

Figure A.20: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 2$, $\rho = 0.75$, $\alpha = 0.50$.

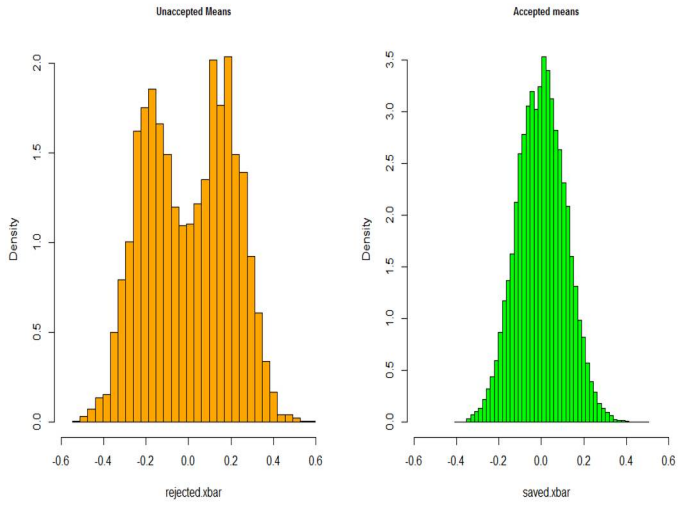

Figure A.21: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 2$, $\rho = 0.75$, $\alpha = 0.75$.

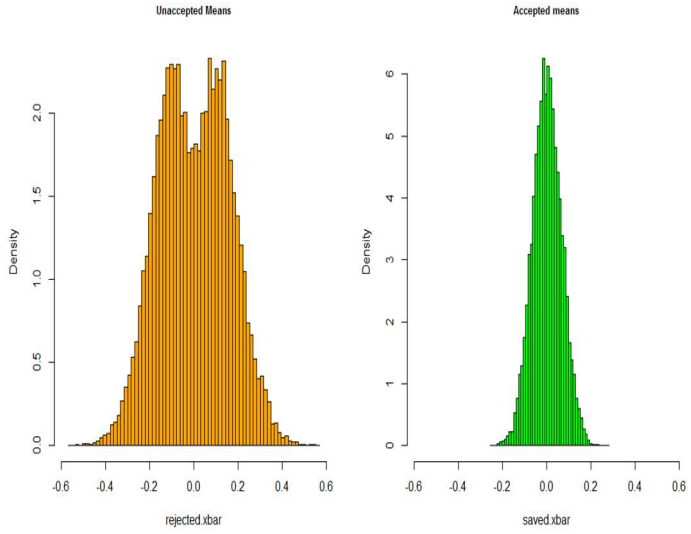

Figure A.22: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 2$, $\rho = 0.90$, $\alpha = 0.25$.

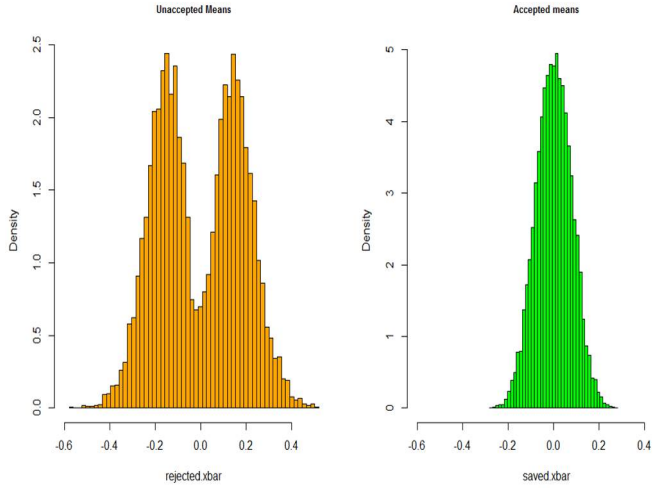

Figure A.23: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 2$, $\rho = 0.90$, $\alpha = 0.50$.

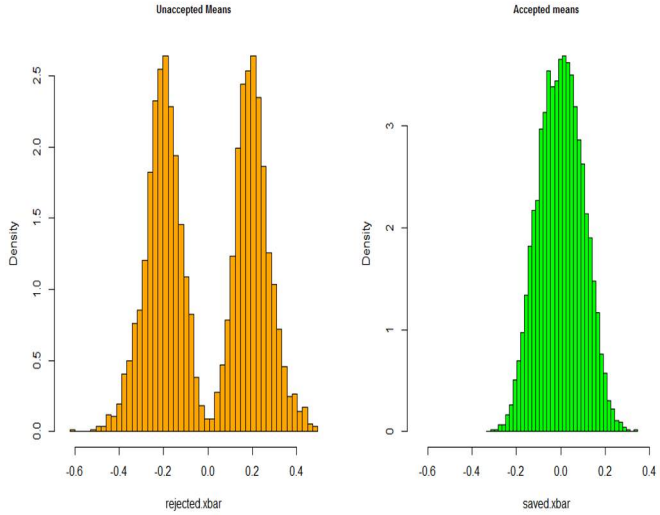

Figure A.24: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 2$, $\rho = 0.90$, $\alpha = 0.75$.

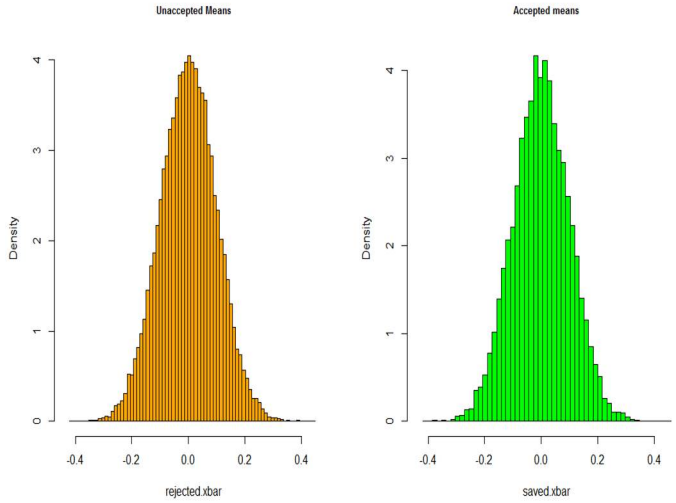

Figure A.25: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 2$, $\rho = 0$, $\alpha = 0.25$.

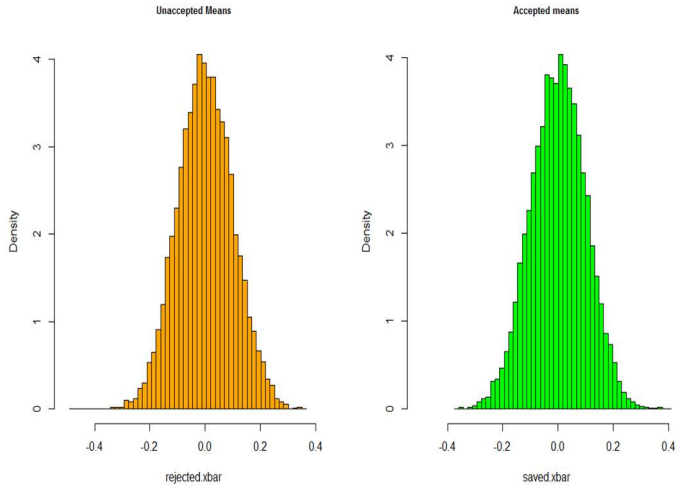

Figure A.26: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 2$, $\rho = 0$, $\alpha = 0.50$.

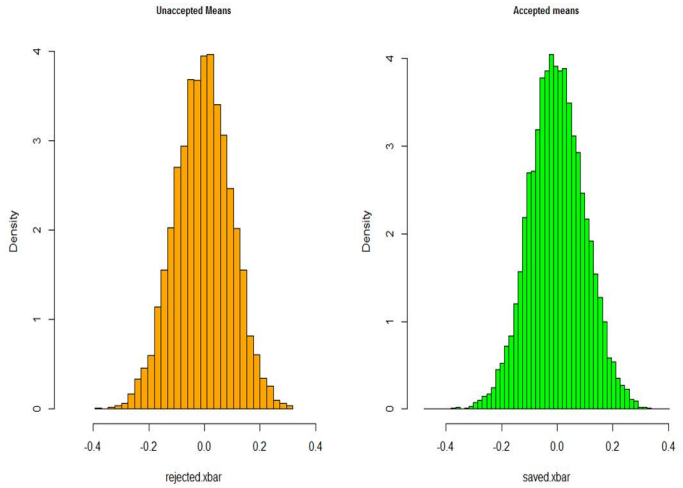

Figure A.27: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 2$, $\rho = 0$, $\alpha = 0.75$.

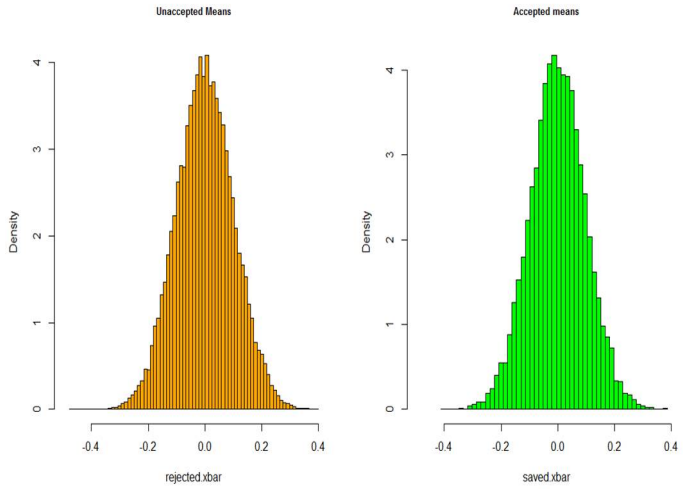

Figure A.28: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 2$, $\rho = 0.25$, $\alpha = 0.25$.

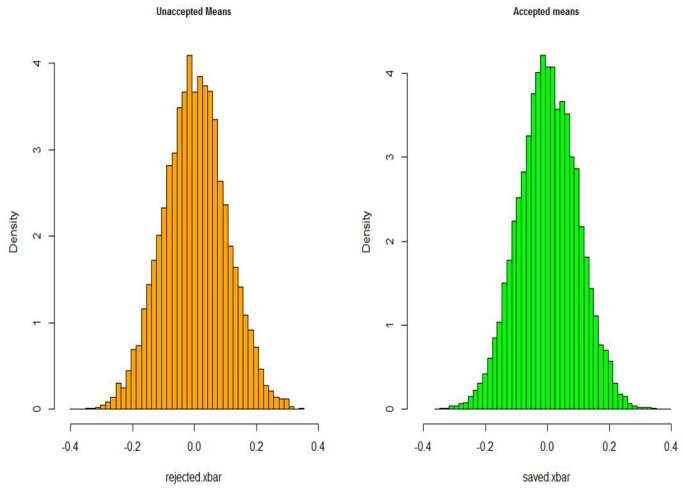

Figure A.29: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 2$, $\rho = 0.25$, $\alpha = 0.50$.

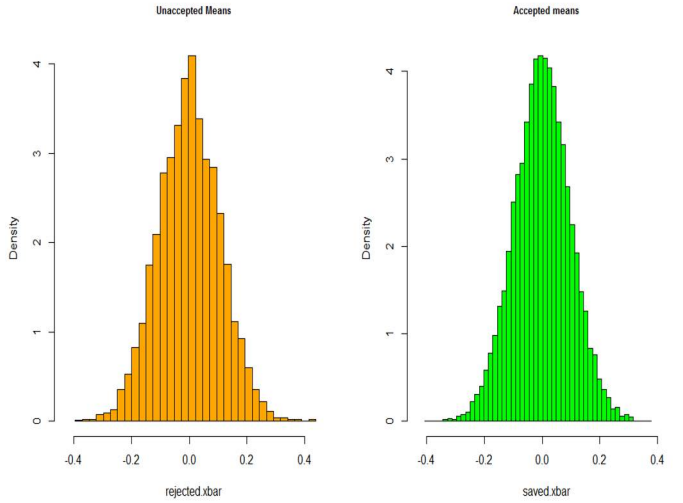

Figure A.30: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 2$, $\rho = 0.25$, $\alpha = 0.75$.

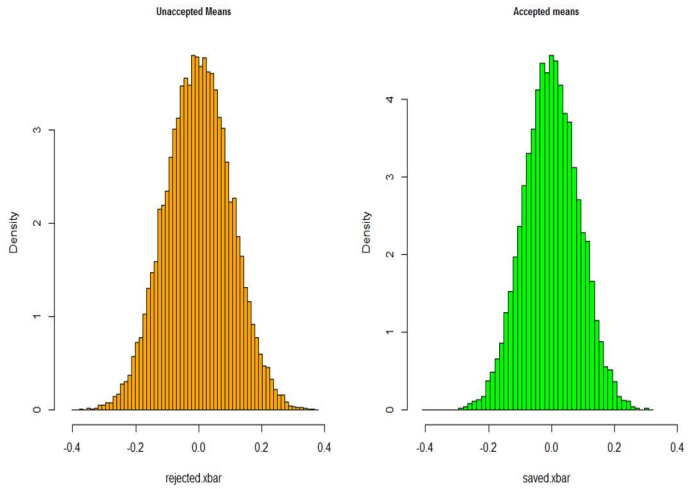

Figure A.31: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 2$, $\rho = 0.50$, $\alpha = 0.25$.

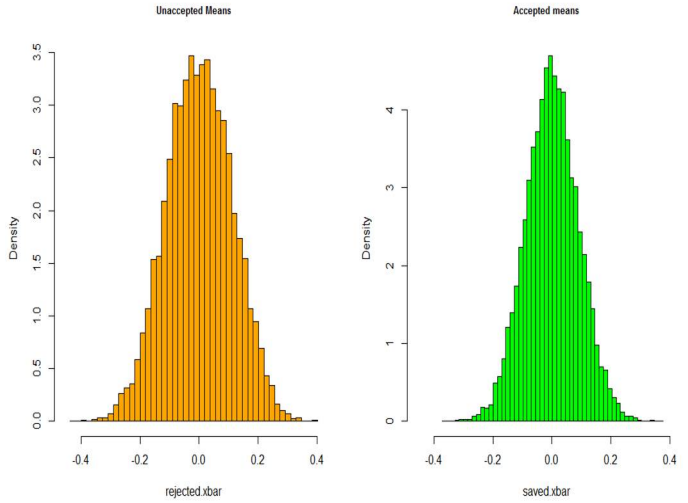

Figure A.32: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 2$, $\rho = 0.50$, $\alpha = 0.50$.

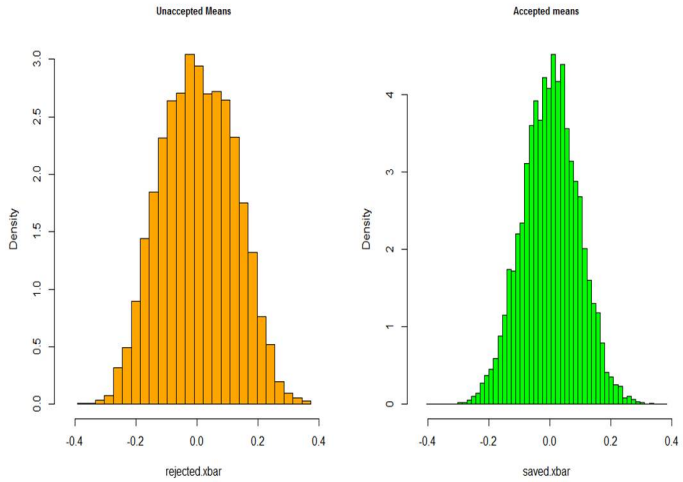

Figure A.33: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 2$, $\rho = 0.50$, $\alpha = 0.75$.

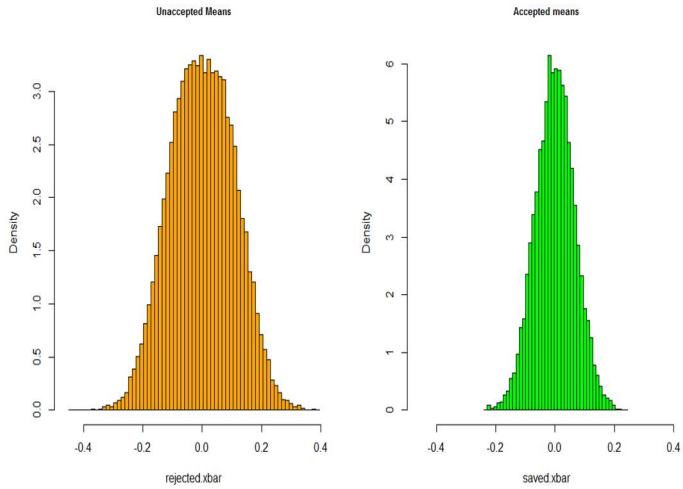

Figure A.34: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 2$, $\rho = 0.75$, $\alpha = 0.25$.

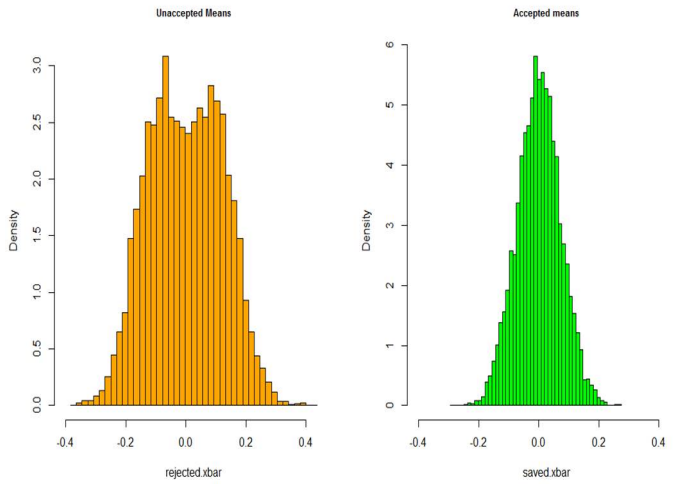

Figure A.35: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 2$, $\rho = 0.75$, $\alpha = 0.50$.

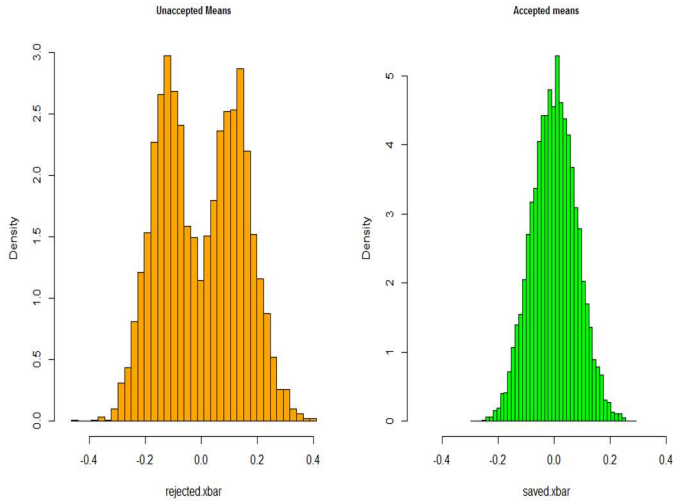

Figure A.36: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 2$, $\rho = 0.75$, $\alpha = 0.75$.

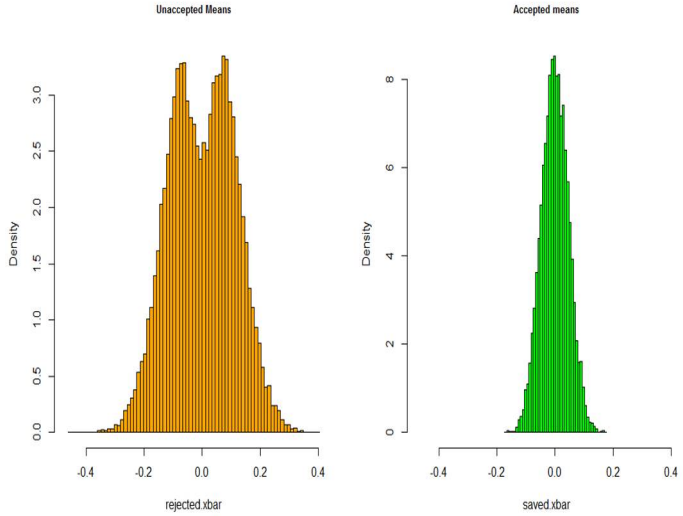

Figure A.37: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 2$, $\rho = 0.90$, $\alpha = 0.25$.

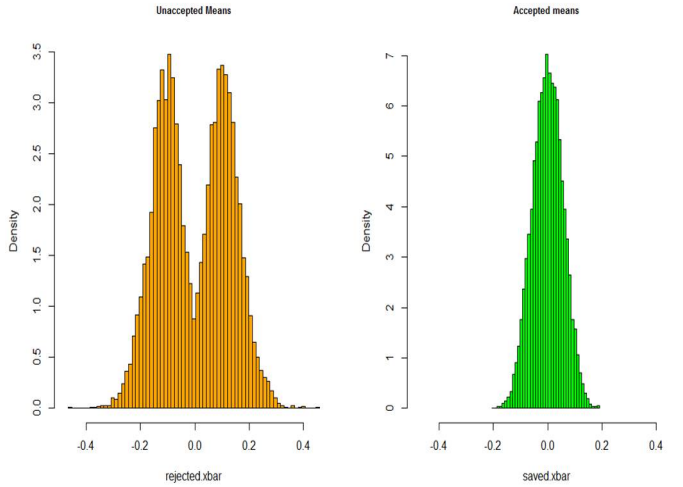

Figure A.38: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 2$, $\rho = 0.90$, $\alpha = 0.50$.

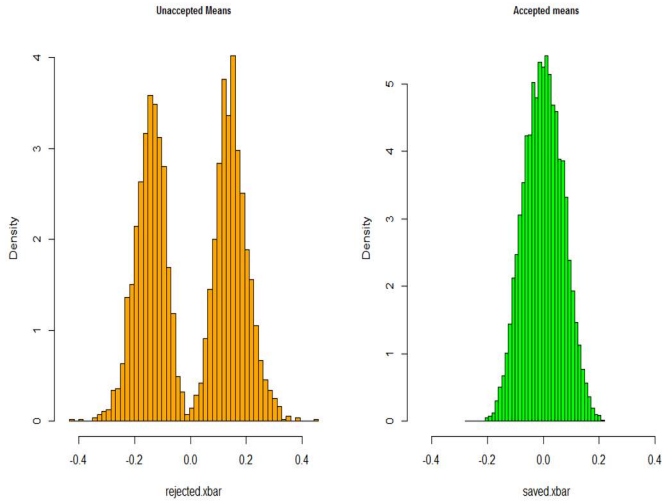

Figure A.39: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 2$, $\rho = 0.90$, $\alpha = 0.75$.

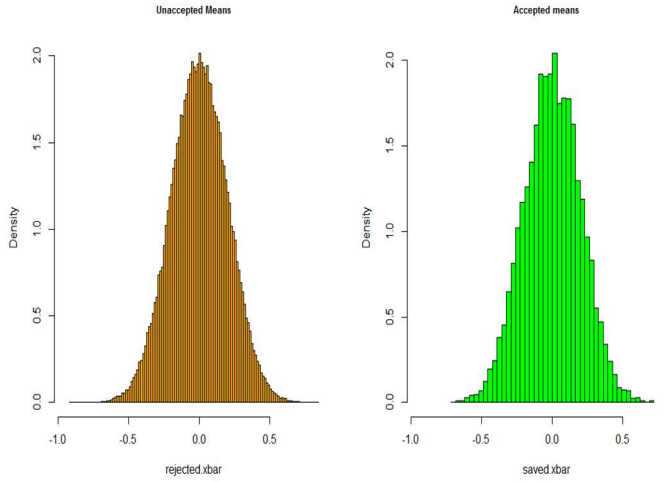

Figure A.40: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 3$, $\rho = 0$, $\alpha = 0.25$.

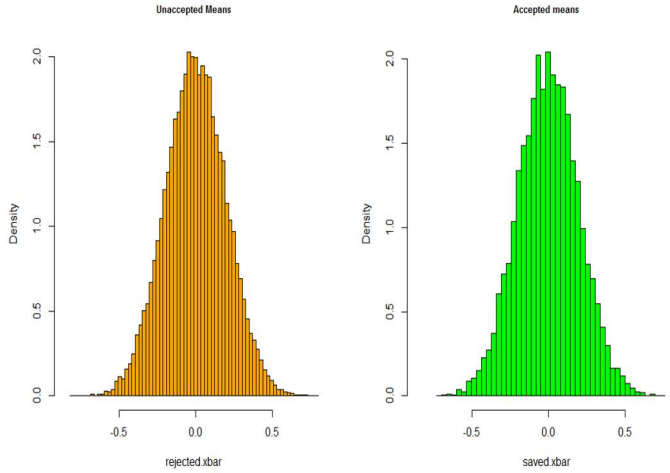

Figure A.41: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 3$, $\rho = 0$, $\alpha = 0.50$.

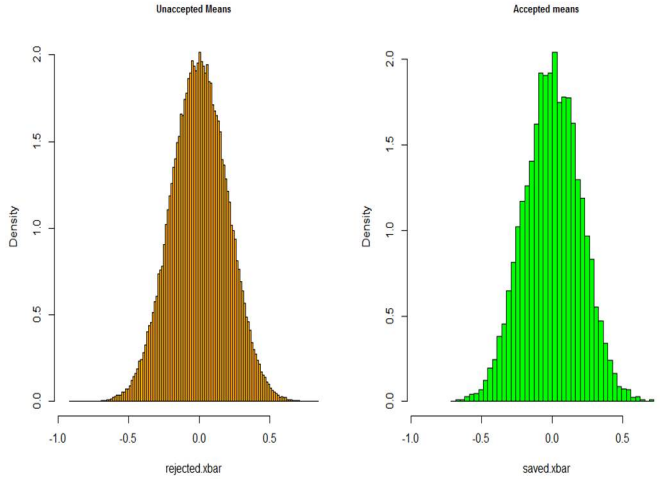

Figure A.42: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 3$, $\rho = 0$, $\alpha = 0.75$.

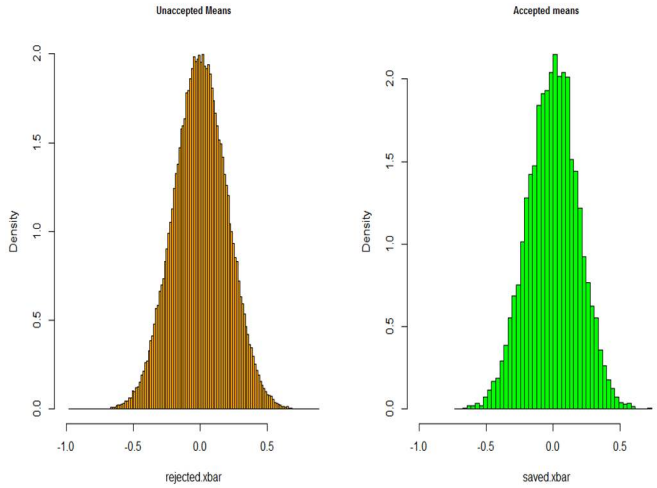

Figure A.43: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 3$, $\rho = 0.25$, $\alpha = 0.25$.

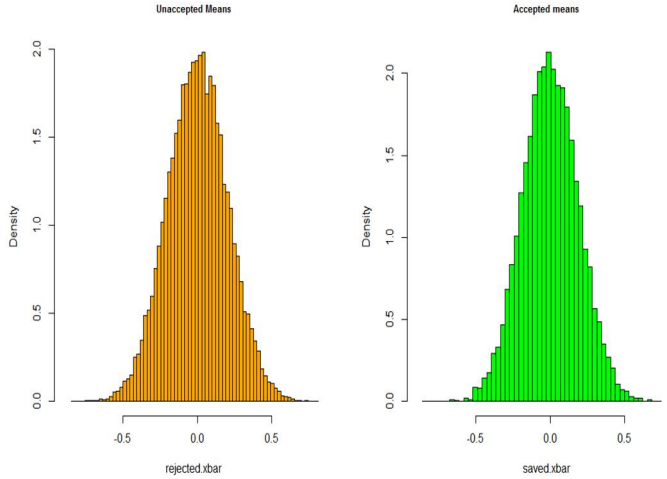

Figure A.44: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 3$, $\rho = 0.25$, $\alpha = 0.50$.

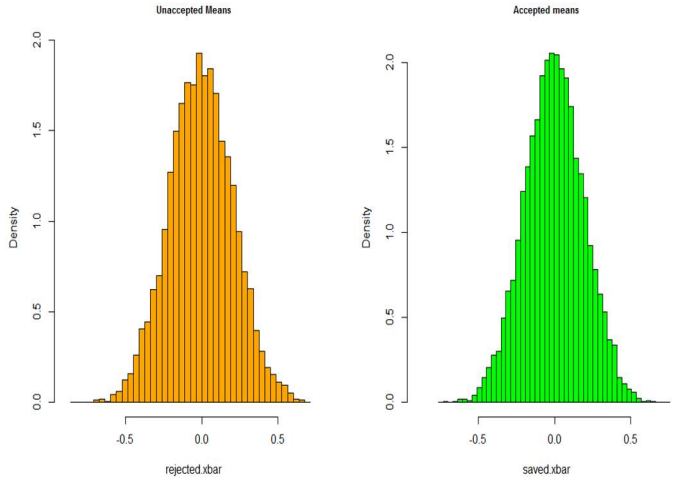

Figure A.45: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 3$, $\rho = 0.25$, $\alpha = 0.75$.

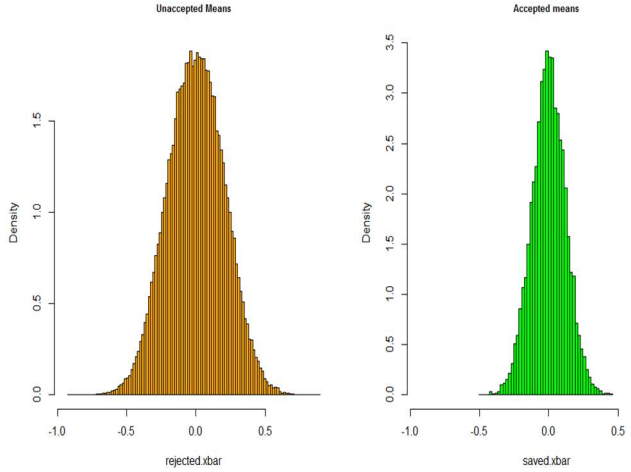

Figure A.46: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 3$, $\rho = 0.75$, $\alpha = 0.25$.

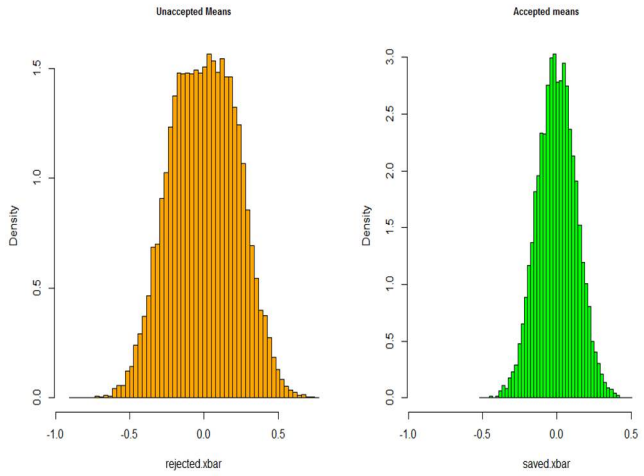

Figure A.47: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 3$, $\rho = 0.75$, $\alpha = 0.50$.

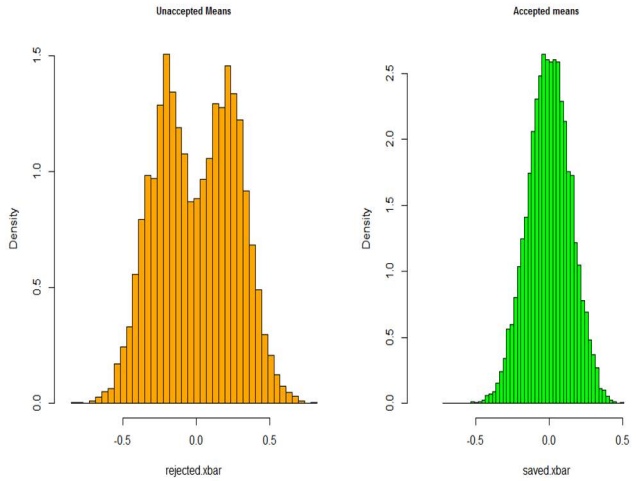

Figure A.48: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 3$, $\rho = 0.75$, $\alpha = 0.75$.

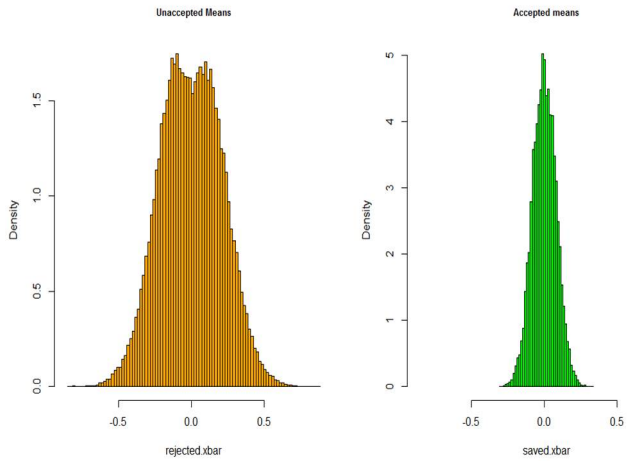

Figure A.49: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 3$, $\rho = 0.90$, $\alpha = 0.25$.

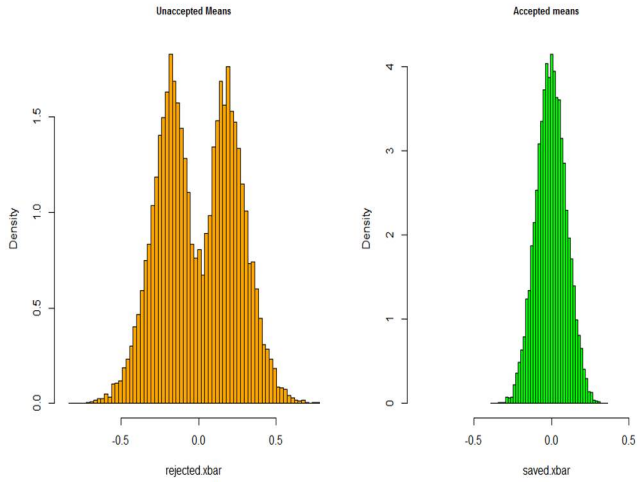

Figure A.50: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 3$, $\rho = 0.90$, $\alpha = 0.50$.

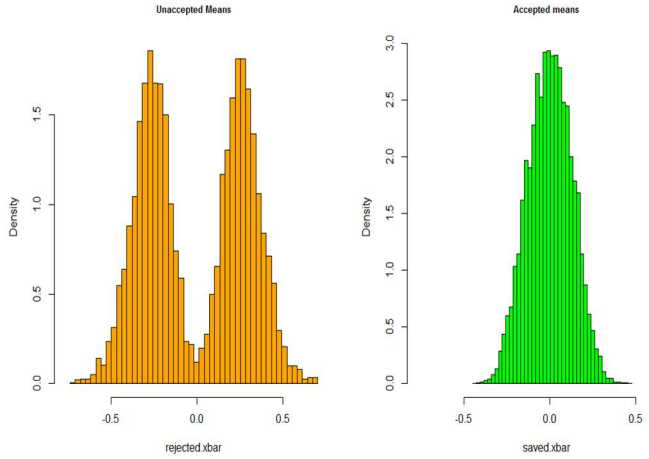

Figure A.51: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 25$, $d = 3$, $\rho = 0.90$, $\alpha = 0.75$.

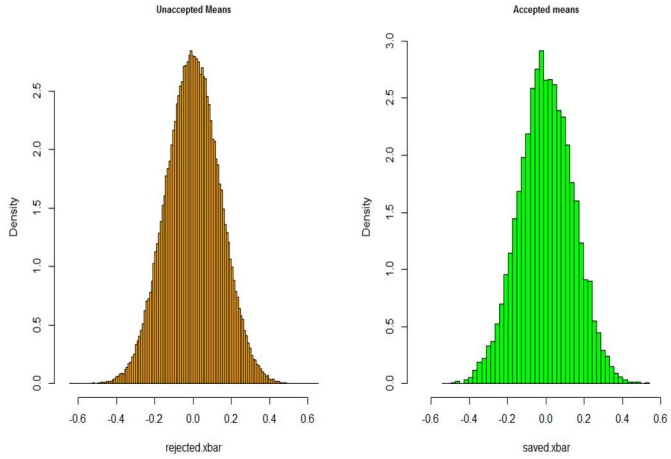

Figure A.52: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 3$, $\rho = 0$, $\alpha = 0.25$.

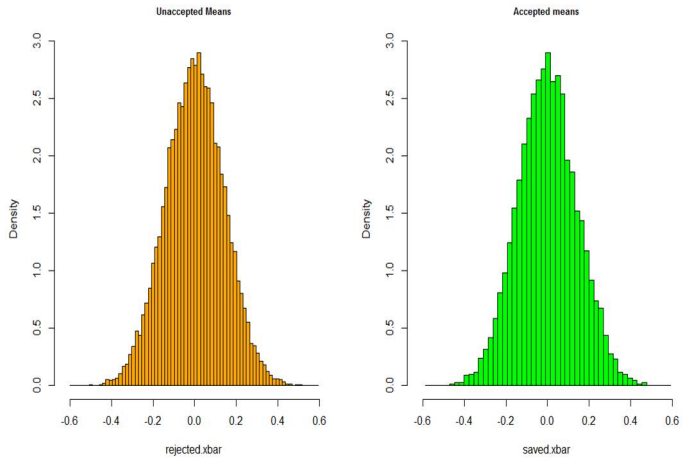

Figure A.53: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 3$, $\rho = 0$, $\alpha = 0.50$.

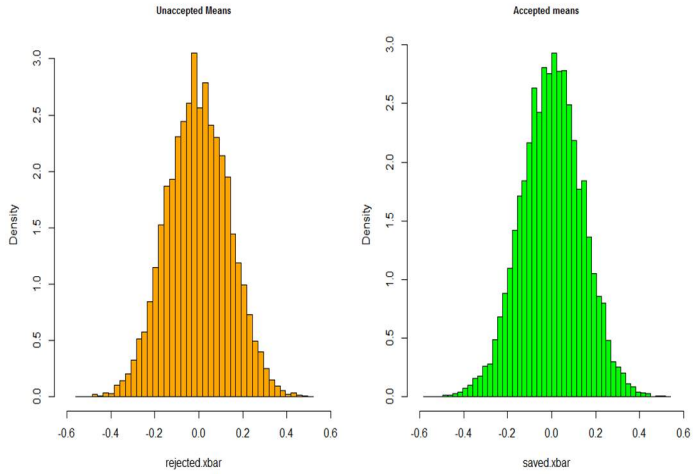

Figure A.54: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 3$, $\rho = 0$, $\alpha = 0.75$.

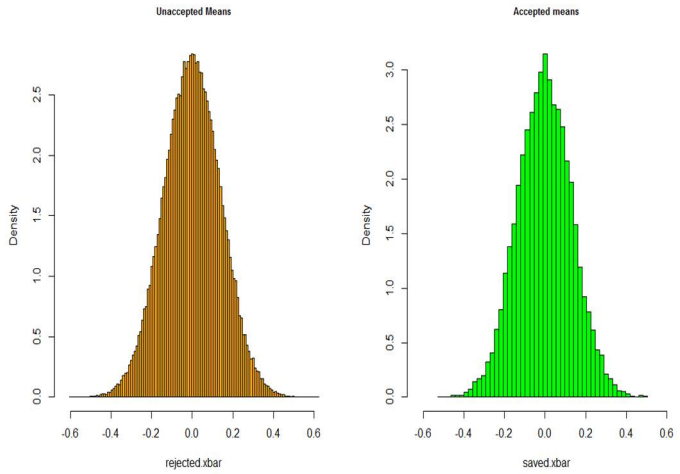

Figure A.55: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 3$, $\rho = 0.25$, $\alpha = 0.25$.

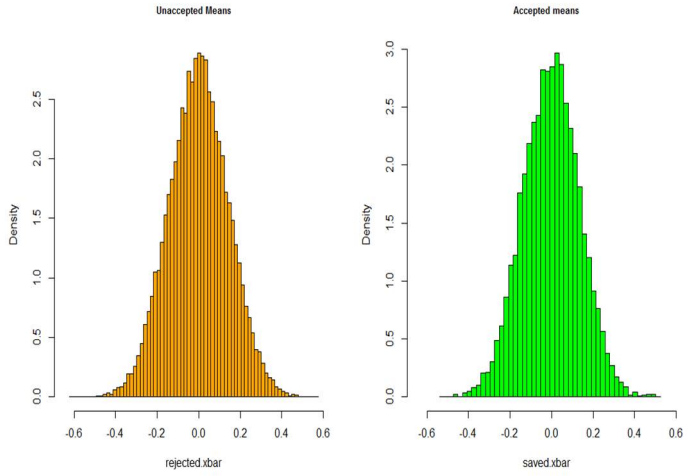

Figure A.56: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 3$, $\rho = 0.25$, $\alpha = 0.50$.

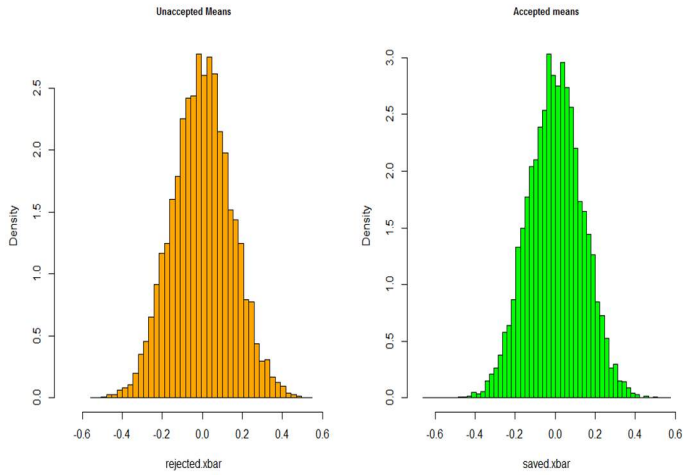

Figure A.57: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 3$, $\rho = 0.25$, $\alpha = 0.75$.

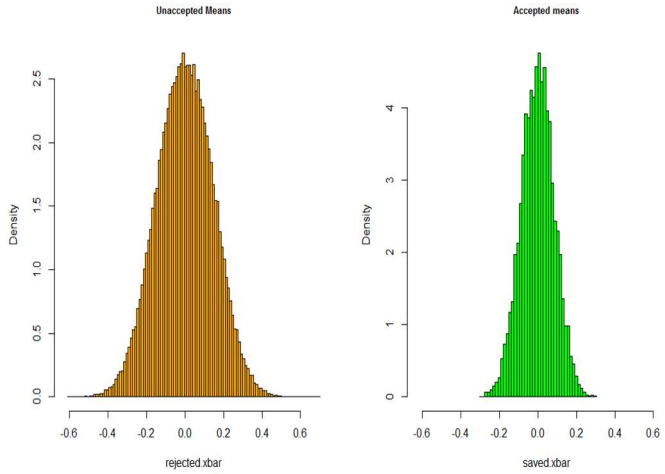

Figure A.58: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 3$, $\rho = 0.75$, $\alpha = 0.25$.

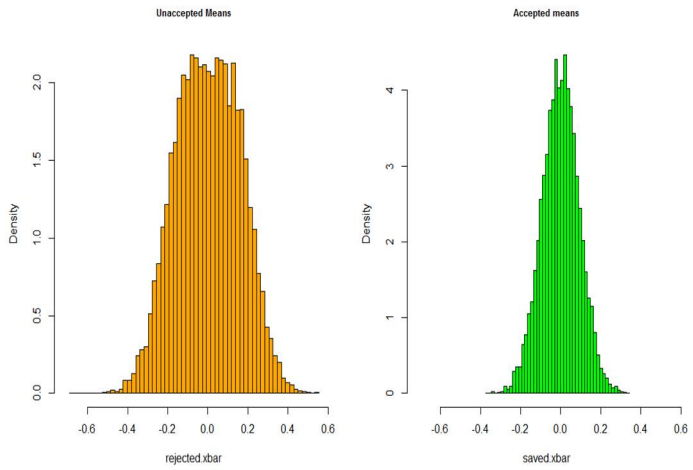

Figure A.59: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 3$, $\rho = 0.75$, $\alpha = 0.50$.

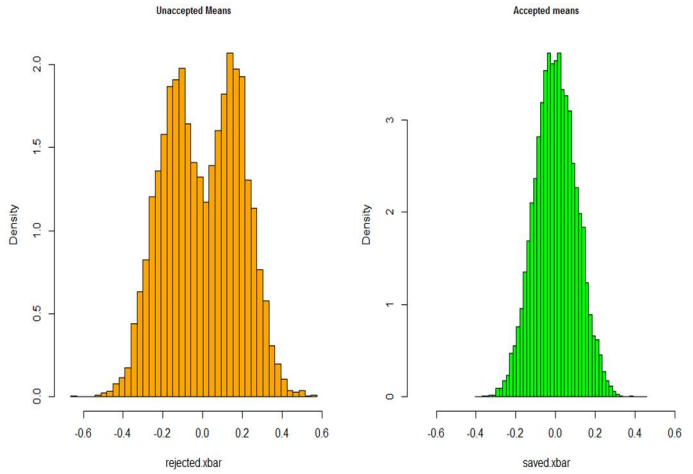

Figure A.60: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 3$, $\rho = 0.75$, $\alpha = 0.75$.

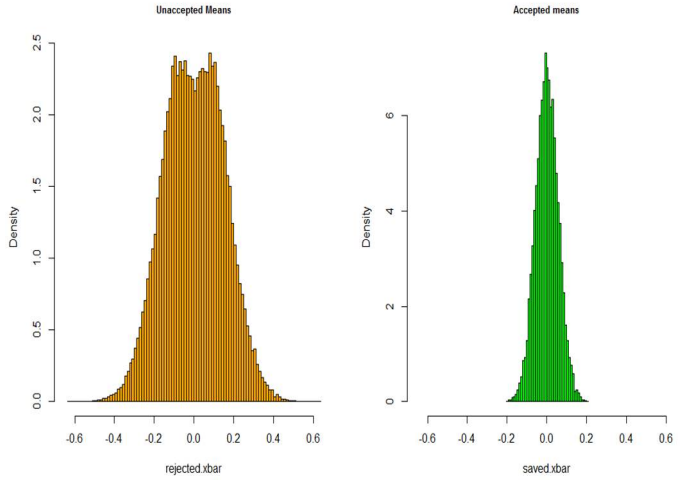

Figure A.61: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 3$, $\rho = 0.90$, $\alpha = 0.25$.

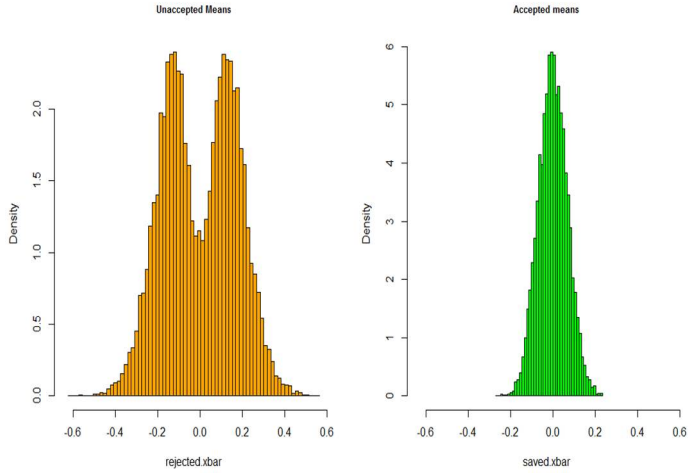

Figure A.62: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 3$, $\rho = 0.90$, $\alpha = 0.50$.

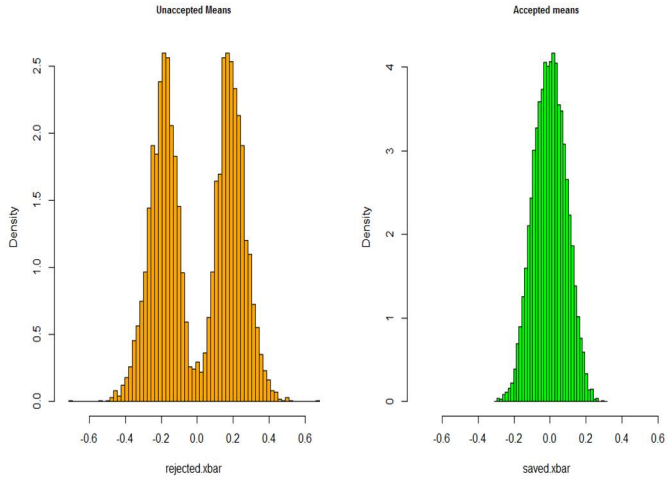

Figure A.63: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 50$, $d = 3$, $\rho = 0.90$, $\alpha = 0.75$.

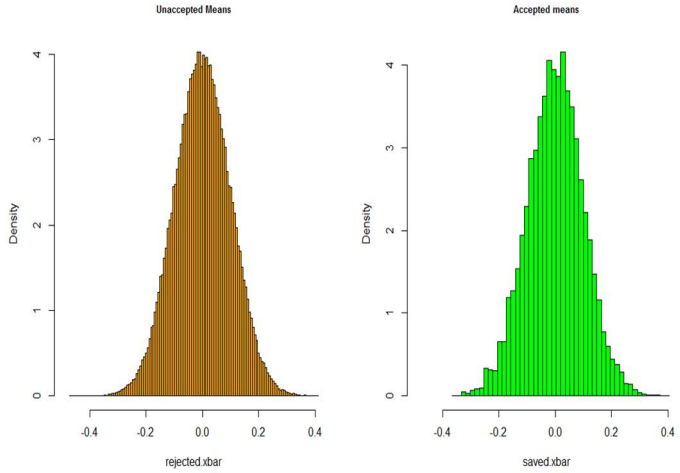

Figure A.64: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 3$, $\rho = 0$, $\alpha = 0.25$.

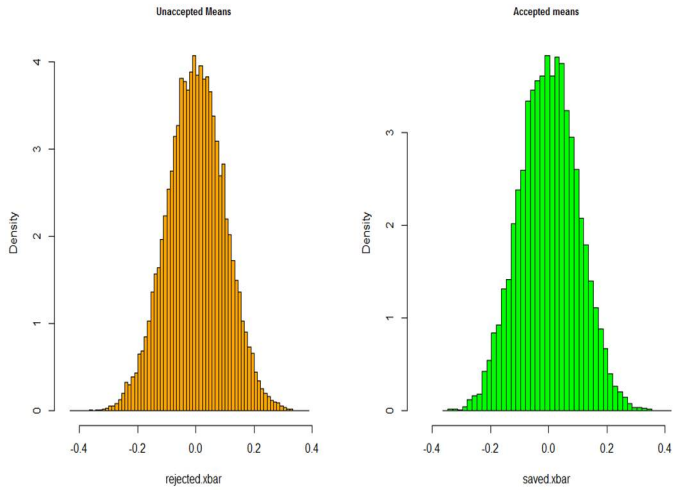

Figure A.65: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 3$, $\rho = 0$, $\alpha = 0.50$.

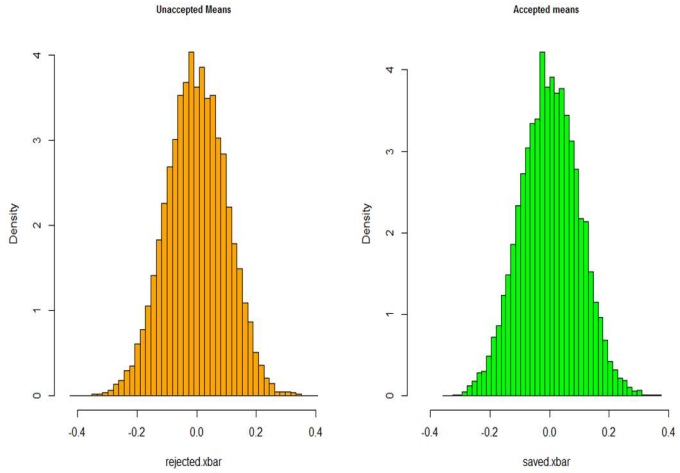

Figure A.66: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 3$, $\rho = 0$, $\alpha = 0.75$.

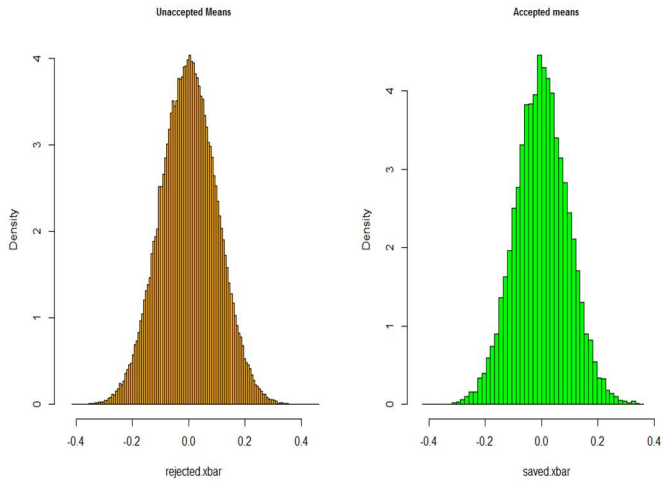

Figure A.67: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 3$, $\rho = 0.25$, $\alpha = 0.25$.

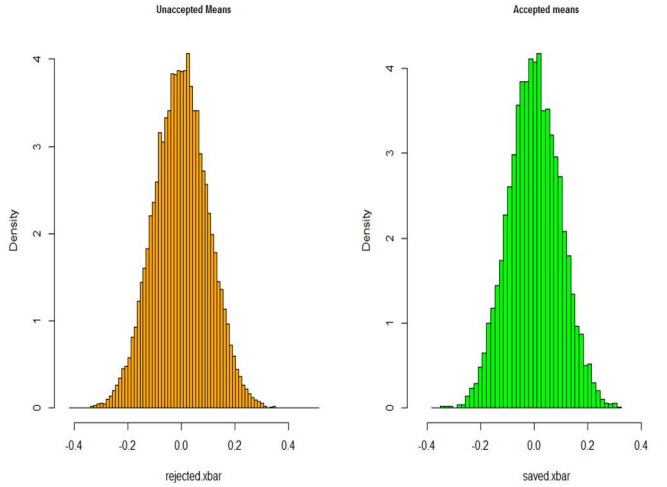

Figure A.68: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 3$, $\rho = 0.25$, $\alpha = 0.50$.

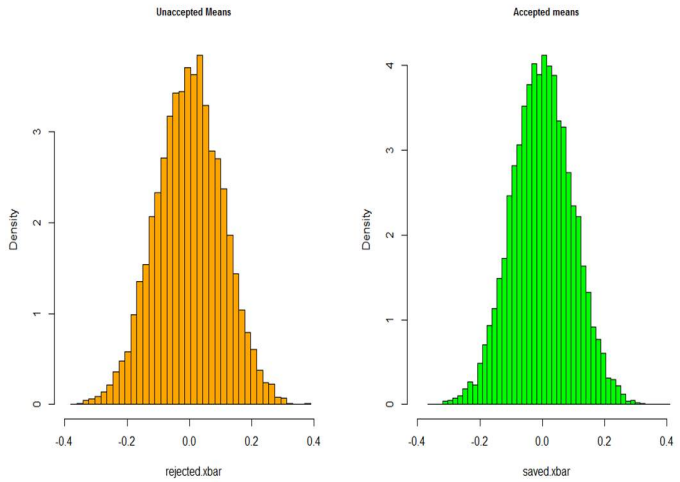

Figure A.69: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 3$, $\rho = 0.25$, $\alpha = 0.75$.

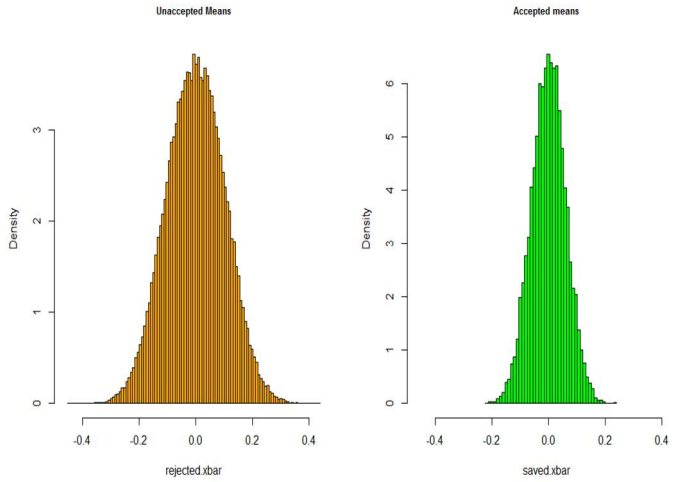

Figure A.70: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 3$, $\rho = 0.75$, $\alpha = 0.25$.

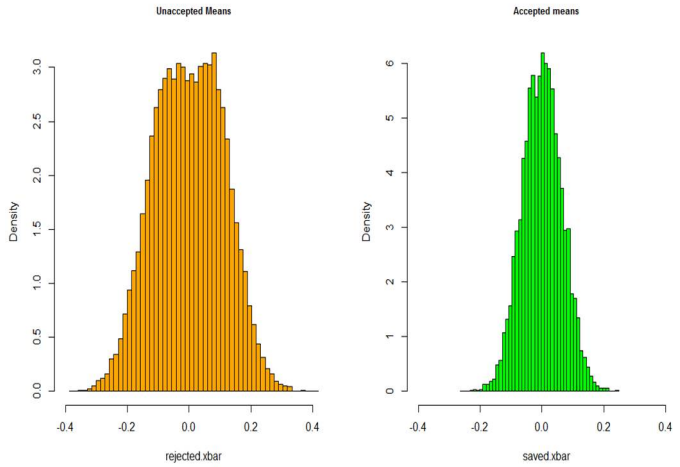

Figure A.71: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 3$, $\rho = 0.75$, $\alpha = 0.50$.

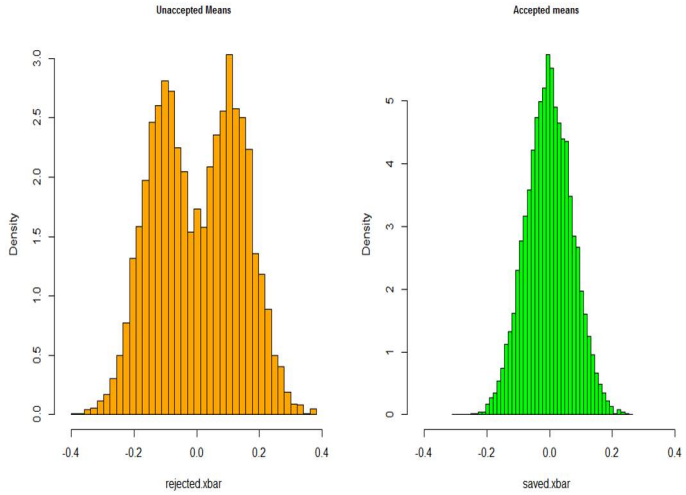

Figure A.72: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 3$, $\rho = 0.75$, $\alpha = 0.75$.

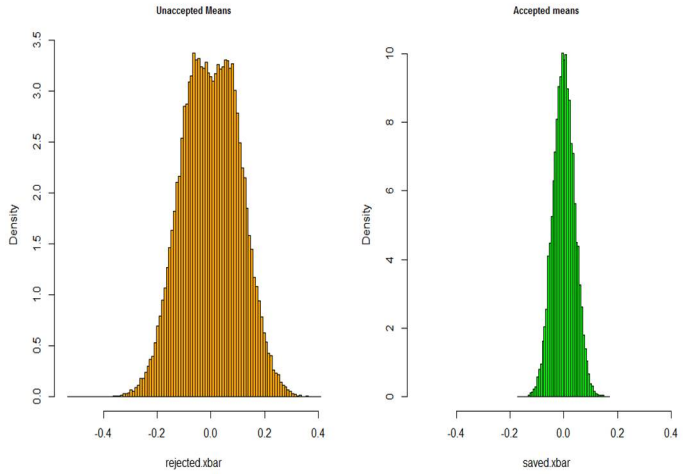

Figure A.73: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 3$, $\rho = 0.90$, $\alpha = 0.25$.

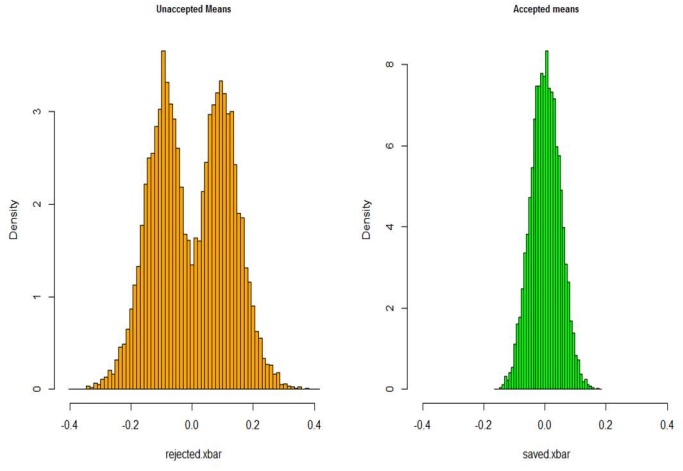

Figure A.74: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 3$, $\rho = 0.90$, $\alpha = 0.50$.

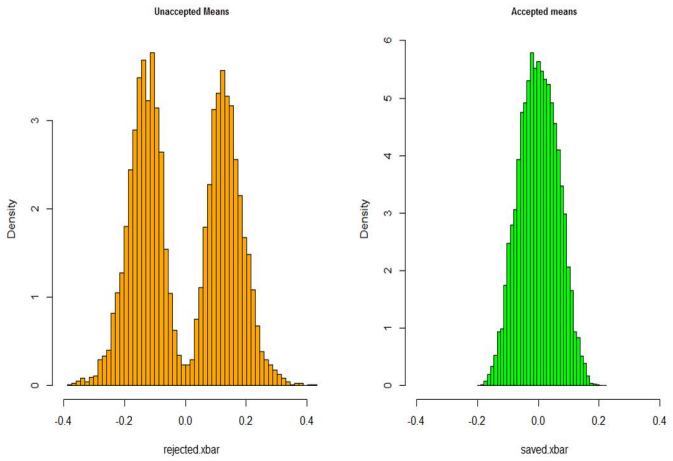

Figure A.75: Histogram of the unaccepted means (left) and the accepted means (right) for $n = 100$, $d = 3$, $\rho = 0.90$, $\alpha = 0.75$.

APPENDIX B

R CODES

R codes we used in our simulation and numerical numerical investigation.

R CODE FOR SIMULATION FOR $d = 2$.

```
library(MASS)
library(KernSmooth)

b <- 10000    # Number of replications
d <- 2       # Dimension
rho <- 0.90     # Correlation
n <- 25   # Sample Size
norm.flag <- 1  # 1 => Normal Samples, 0 <= Non-normal Samples

#
#  Set up mean vector correlation matrix for Normal samples
#

m <- rep(0,d)
I <- diag(rep(1,d))
J <- array(1,dim=c(d,d))
sigma <- ((1-rho)*I) + (rho*J)

#
#  Function for Generating Non-Normal Samples
#

three.mix <- function(n) {
```

```
p <- sample(1:3,n,replace=T)
S <- array(dim=c(2,2,3))
S[,,1] <- matrix(c(1,0,0,5),2,2)
S[,,2] <- matrix(c(1,-0.6,-0.6,1),2,2)
S[,,3] <- matrix(c(5,0,0,1),2,2)
m <- array(dim=c(2,3))
m[,1] <- c(1,7)
m[,2] <- c(1.9,1.9)
m[,3] <- c(7,1)
x <- matrix(0,n,2)
for(i in 1:n) x[i,] <- mvrnorm(1,m[,p[i]],S[,,p[i]])-c(3.3,3.3)
return(x)
}

#
#  Counter for the number of means that have been saved
#

counter <- 1

#
# Vector of saved means
#

saved.xbar <- matrix(0,b,1)
rejected.xbar <- c()
```

```
#
# Main simulation loop
#

while(counter<=b) {

#
#   Simulates a sample of size n, dimension d
# and extracts the first column
#

if(norm.flag==0) X <- three.mix(n)
if(norm.flag==1) X <- mvrnorm(n,m,sigma)
if(norm.flag==2) X <- nsm(n)

#
#  Computes the mean of the first column
#

X1 <- apply(X,2,mean)
#
#   Flag for detecting if the sample is accepted
#

flag <- 1
```

```
#
# Checks to see if a confidence interval for the mean of each column
# contains the true mean. If not, the sample is rejected.
#

for(i in 2:d) {
     ci <- t.test(X[,i],conf.level=0.99)$conf.int
     if((ci[1]>0)||(ci[2]<0)) flag <- 0
   }
   if(flag==0) {
     cat(paste(counter," Rejecting...\n"))
     rejected.xbar <- c(rejected.xbar,X1[1])
   }
#
# If the sample is not rejected, then save the mean
#

if(flag==1) {
     saved.xbar[counter] <- X1[1]
     counter <- counter + 1
}}

#
# Compute theoretical standard error for sample mean and the
# observed standard error for the accepted means
```

```
#

std.error <- 1/sqrt(n)
bias <- 0
mse <- std.error
val.std.error <- sqrt(sum(saved.xbar^2)/(b-1))
val.bias <- mean(saved.xbar)
val.mse <- sqrt(val.std.error^2+val.bias^2)

PA <- length(saved.xbar)/(length(saved.xbar)+length(rejected.xbar))

cat("-----------------------------------------------------\n")
cat(" Simulation Results, Acceptance Sampling Methodology  \n")
cat("-----------------------------------------------------\n")
cat(" Sample Size                              ",n,"\n")
cat(" Dimension                                ",d,"\n")
cat(" Correlation                              ",rho,"\n")
cat(" Distribution                              ")
if(norm.flag==1) cat("Normal\n")
if(norm.flag==0) cat("Normal Mixture\n")
cat("-----------------------------------------------------\n")
cat(" Non-validated Standard Error           ",std.error," \n")
cat(" Non-validated Bias                     ",bias,"\n")
cat(" Non-validated Root Mean Squared Error  ",mse,"\n")
cat("-----------------------------------------------------\n")
cat(" Validated Standard Error               ",val.std.error," \n")
```

```
cat(" Validated Bias                          ",val.bias,"\n")
cat(" Validated Root Mean Squared Error       ",val.mse,"\n")
cat(" Proportion of Means Accepted            ",PA,"\n")
cat("-----------------------------------------------------\n")

#
#  Plot the Histograms
#
#
#  Sets up a common range for the x-axis of the two histograms
# and compute bin sizes based on the Wand algorithm
#

par(mfrow=c(1,2))
h1 <- dpih(rejected.xbar)
bins1 <- seq(min(rejected.xbar)-h1, max(rejected.xbar)+h1, by=h1)
h2 <- dpih(saved.xbar)
bins2 <- seq(min(saved.xbar)-h2, max(saved.xbar)+h2, by=h2)

xl <- c(min(bins1,bins2)-max(h1,h2),max(bins1,bins2)+max(h1,h2))

colors = c("range","green")

hist(rejected.xbar,prob=T,main="Unaccepted Means",freq=F,cex.main=0.8,
     xlim=xl,breaks=bins1, col = "orange")
hist(saved.xbar,prob=T,main="Accepted means",freq=F,cex.main=0.8,xlim=xl,
```

```
     breaks=bins2, col = "green")
par(mfrow=c(1,1))
```

R CODE FOR SIMULATION FOR $d = 3$.

```
library(MASS)
#library(car)
#library(ellipse)
library(KernSmooth)

b <- 10000    # Number of replications
d <- 3       # Dimension
rho <- 0.90     # Correlation
n <- 100   # Sample Size
norm.flag <- 1  # 1 => Normal Samples, 0 <= Non-normal Samples

#
#  Set up mean vector correlation matrix for Normal samples
#

m <- rep(0,d)
I <- diag(rep(1,d))
J <- array(1,dim=c(d,d))
sigma <- ((1-rho)*I) + (rho*J)

#
```

```
#  Function for Generating Non-Normal Samples
#

three.mix <- function(n) {
p <- sample(1:3,n,replace=T)
S <- array(dim=c(3,3,3))
S[,,1] <- matrix(c(1,0,0,0,1,0,0,0,5),3,3)
S[,,2] <- matrix(c(1,-0.6,0,-0.6,1,0,-0.6,0,1),3,3)
S[,,3] <- matrix(c(5,0,0,0,1,0,0,0,1),3,3)
m <- array(dim=c(3,3))
m[,1] <- c(1,1,7)
m[,2] <- c(1.9,1.9,1.9)
m[,3] <- c(7,1,1)
x <- matrix(0,n,3)
for(i in 1:n) x[i,] <- mvrnorm(1,m[,p[i]],S[,,p[i]])-c(3.3,1.3,3.3)
return(x)
}

#
#  Counter for the number of means that have been saved
#

counter <- 1

#
# Vector of saved means
```

```
#

saved.xbar <- matrix(0,b,1)
rejected.xbar <- c()

#
# Main simulation loop
#

while(counter<=b) {

#
#   Simulates a sample of size n, dimension d
# and extracts the first column
#

if(norm.flag==0) X <- three.mix(n)
if(norm.flag==1) X <- mvrnorm(n,m,sigma)
if(norm.flag==2) X <- nsm(n)

#
#  Computes the mean of the first column
#

X1 <- apply(X,2,mean)
#
```

```
#   Flag for detecting if the sample is accepted
#

flag <- 1

#
# Checks to see if a confidence interval for the mean of each column
# contains the true mean. If not, the sample is rejected.
#

for(i in 2:d) {
     ci <- t.test(X[,i],conf.level=0.75)$conf.int
     if((ci[1]>0)||(ci[2]<0)) flag <- 0
   }
   if(flag==0) {
     cat(paste(counter," Rejecting...\n"))
     rejected.xbar <- c(rejected.xbar,X1[1])
   }
#
# If the sample is not rejected, then save the mean
#

if(flag==1) {
     saved.xbar[counter] <- X1[1]
     counter <- counter + 1
}}
```

```
#
# Compute theoretical standard error for sample mean and the
# observed standard error for the accepted means
#

std.error <- 1/sqrt(n)
bias <- 0
mse <- std.error
val.std.error <- sqrt(sum(saved.xbar^2)/(b-1))
val.bias <- mean(saved.xbar)
val.mse <- sqrt(val.std.error^2+val.bias^2)

PA <- length(saved.xbar)/(length(saved.xbar)+length(rejected.xbar))

cat("-----------------------------------------------------\n")
cat(" Simulation Results, Acceptance Sampling Methodology  \n")
cat("-----------------------------------------------------\n")
cat(" Sample Size                              ",n,"\n")
cat(" Dimension                                ",d,"\n")
cat(" Correlation                              ",rho,"\n")
cat(" Distribution                              ")
if(norm.flag==1) cat("Normal\n")
if(norm.flag==0) cat("Normal Mixture\n")
cat("-----------------------------------------------------\n")
cat(" Non-validated Standard Error           ",std.error," \n")
```

```
cat(" Non-validated Bias                      ",bias,"\n")
cat(" Non-validated Root Mean Squared Error  ",mse,"\n")
cat("-----------------------------------------------------\n")
cat(" Validated Standard Error               ",val.std.error," \n")
cat(" Validated Bias                         ",val.bias,"\n")
cat(" Validated Root Mean Squared Error      ",val.mse,"\n")
cat(" Proportion of Means Accepted           ",PA,"\n")
cat(" Confidence Interval           ",ci,"\n")
cat("-----------------------------------------------------\n")

#
#  Plot the Histograms
#
#
#  Sets up a common range for the x-axis of the two histograms
# and compute bin sizes based on the Wand algorithm
#

par(mfrow=c(1,2))
h1 <- dpih(rejected.xbar)
bins1 <- seq(min(rejected.xbar)-h1, max(rejected.xbar)+h1, by=h1)
h2 <- dpih(saved.xbar)
bins2 <- seq(min(saved.xbar)-h2, max(saved.xbar)+h2, by=h2)

xl <- c(min(bins1,bins2)-max(h1,h2),max(bins1,bins2)+max(h1,h2))
```

```
colors = c("range","green")

hist(rejected.xbar,prob=T,main="Unaccepted Means",freq=F,cex.main=0.8,
     xlim=xl,breaks=bins1, col = "orange")
hist(saved.xbar,prob=T,main="Accepted means",freq=F,cex.main=0.8,xlim=xl,
     breaks=bins2, col = "green")
par(mfrow=c(1,1))
```

R CODES SIMULATION USING CONTROL VARIATES

R CODE FOR SIMULATION USING CONTROL VARIATE FOR $d = 2$.

```
library(MASS)

b <- 10000    # Number of replications
d <- 2       # Dimension
rho <- 0.0     # Correlation
n <- 25   # Sample Size
norm.flag <- 1  # 1 => Normal Samples, 0 <= Non-normal Samples

#
#  Set up mean vector correlation matrix for Normal samples
#

m <- rep(0,d)
I <- diag(rep(1,d))
```

```
J <- array(1,dim=c(d,d))
sigma <- ((1-rho)*I) + (rho*J)

#
#  Function for Generating Non-Normal Samples
#

three.mix <- function(n) {
p <- sample(1:3,n,replace=T)
S <- array(dim=c(2,2,3))
S[,,1] <- matrix(c(1,0,0,5),2,2)
S[,,2] <- matrix(c(1,-0.6,-0.6,1),2,2)
S[,,3] <- matrix(c(5,0,0,1),2,2)
m <- array(dim=c(2,3))
m[,1] <- c(1,7)
m[,2] <- c(1.9,1.9)
m[,3] <- c(7,1)
x <- matrix(0,n,2)
for(i in 1:n) x[i,] <- mvrnorm(1,m[,p[i]],S[,,p[i]])-c(3.3,3.3)
return(x)
}

#
# Main simulation loop
#
```

```
adj.xbar <- matrix(0,b,1)

for(i in 1:b) {

#
#   Simulates a sample of size n, dimension d
# and extracts the first column
#

if(norm.flag==0) X <- three.mix(n)
if(norm.flag==1) X <- mvrnorm(n,m,sigma)

#
# Estimate the optimal coefficient and adjust the mean
#

M <- apply(X,2,mean)
V <- var(X)
adj.xbar[i] <- M[1]-V[1,2]/V[2,2]*M[2]
}

#
# Compute theoretical standard error for sample mean and the
# observed standard error for the accepted means
#
```

```
std.error <- 1/sqrt(n)

bias <- 0

mse <- std.error

adj.std.error <- sqrt(sum(adj.xbar^2)/(b-1))

adj.bias <- mean(adj.xbar)

adj.mse <- sqrt(adj.std.error^2+adj.bias^2)


cat("-----------------------------------------------------\n")

cat(" Simulation Results, Control Variate Methodology  \n")

cat("-----------------------------------------------------\n")

cat(" Sample Size                              ",n,"\n")

cat(" Dimension                                ",d,"\n")

cat(" Correlation                              ",rho,"\n")

cat(" Distribution                              ")

if(norm.flag==1) cat("Normal\n")

if(norm.flag==0) cat("Normal Mixture\n")

cat("-----------------------------------------------------\n")

cat(" Non-adjusted  Standard Error           ",std.error," \n")

cat(" Non-adjusted  Bias                     ",bias,"\n")

cat(" Non-adjusted  Root Mean Squared Error  ",mse,"\n")

cat("-----------------------------------------------------\n")

cat(" Adjusted  Standard Error               ",adj.std.error," \n")

cat(" Adjusted Bias                         ",adj.bias,"\n")

cat(" Adjusted Root Mean Squared Error      ",adj.mse,"\n")

cat("-----------------------------------------------------\n")
```

R CODE FOR SIMULATION USING CONTROL VARIATE FOR $d = 3$.

```
library(MASS)

b <- 10000    # Number of replications
d <- 3       # Dimension
rho <- 0.0     # Correlation
n <- 25   # Sample Size
norm.flag <- 1  # 1 => Normal Samples, 0 <= Non-normal Samples

#
#  Set up mean vector correlation matrix for Normal samples
#

m <- rep(0,d)
I <- diag(rep(1,d))
J <- array(1,dim=c(d,d))
sigma <- ((1-rho)*I) + (rho*J)

#
#  Function for Generating Non-Normal Samples
#
  three.mix <- function(n){
    p <- sample(1:3,n,replace=T)
    S <- array(dim=c(3,3,3))
```

```
    S[,,1] <- matrix(c(1,0,0,0,1,0,0,0,5),3,3)
    S[,,2] <- matrix(c(1,-0.6,0,0,0, -0.6,0,0,1,0),3,3)
    S[,,3] <- matrix(c(5,0,0,0,1,0,0,0,1),3,3)
    m <- array(dim=c(3,3))
    m[,1] <- c(1,1,7)
    m[,2] <- c(1.9,1.9,1.9)
    m[,3] <- c(7,1,1)
    x <- matrix(0,n,3)
    for(i in 1:n) x[i,] <- mvrnorm(1,m[,p[i]],S[,,p[i]])-c(3.3,1.3,3.3)
    return(x)
}

#
# Main simulation loop
#

adj.xbar <- matrix(0,b,1)

for(i in 1:b) {

#
#   Simulates a sample of size n, dimension d
# and extracts the first column
#

if(norm.flag==0) X <- three.mix(n)
```

```
if(norm.flag==1) X <- mvrnorm(n,m,sigma)

#
# Estimate the optimal coefficient and adjust the mean
#

M <- apply(X,2,mean)
V <- var(X)
adj.xbar[i] <- M[1]-V[1,2]/V[2,2]*M[2]
}

#
# Compute theoretical standard error for sample mean and the
# observed standard error for the accepted means
#

std.error <- 1/sqrt(n)
bias <- 0
mse <- std.error
adj.std.error <- sqrt(sum(adj.xbar^2)/(b-1))
adj.bias <- mean(adj.xbar)
adj.mse <- sqrt(adj.std.error^2+adj.bias^2)

cat("-----------------------------------------------------\n")
cat(" Simulation Results, Control Variate Methodology  \n")
cat("-----------------------------------------------------\n")
```

```
cat(" Sample Size                              ",n,"\n")
cat(" Dimension                                ",d,"\n")
cat(" Correlation                              ",rho,"\n")
cat(" Distribution                              ")
if(norm.flag==1) cat("Normal\n")
if(norm.flag==0) cat("Normal Mixture\n")
cat("-----------------------------------------------------\n")
cat(" Non-adjusted  Standard Error           ",std.error," \n")
cat(" Non-adjusted  Bias                     ",bias,"\n")
cat(" Non-adjusted  Root Mean Squared Error  ",mse,"\n")
cat("-----------------------------------------------------\n")
cat(" Adjusted  Standard Error               ",adj.std.error," \n")
cat(" Adjusted Bias                         ",adj.bias,"\n")
cat(" Adjusted Root Mean Squared Error      ",adj.mse,"\n")
cat("-----------------------------------------------------\n")
```